

University of Windsor

## Scholarship at UWindor

---

Electronic Theses and Dissertations

Theses, Dissertations, and Major Papers

---

1-1-2007

### Content-based image retrieval using Generic Fourier Descriptor and Gabor Filters.

Quan He

*University of Windsor*

Follow this and additional works at: <https://scholar.uwindsor.ca/etd>

---

#### Recommended Citation

He, Quan, "Content-based image retrieval using Generic Fourier Descriptor and Gabor Filters." (2007).  
*Electronic Theses and Dissertations*. 7046.  
<https://scholar.uwindsor.ca/etd/7046>

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email ([scholarship@uwindsor.ca](mailto:scholarship@uwindsor.ca)) or by telephone at 519-253-3000ext. 3208.

**CONTENT-BASED IMAGE RETRIEVAL USING GENERIC FOURIER  
DESCRIPTOR AND GABOR FILTERS**

by

**Quan He**

**A Thesis**

**Submitted to the Faculty of Graduate Studies  
through the Department of Electrical and Computer Engineering  
in Partial Fulfillment of the Requirements for  
the Degree of Master of Applied Science at the  
University of Windsor**

**Windsor, Ontario, Canada**

**©2007 Quan He**



Library and  
Archives Canada

Bibliothèque et  
Archives Canada

Published Heritage  
Branch

Direction du  
Patrimoine de l'édition

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file    Votre référence*

*ISBN: 978-0-494-35177-2*

*Our file    Notre référence*

*ISBN: 978-0-494-35177-2*

#### NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

#### AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

  
**Canada**

## ABSTRACT

Content-based image retrieval (CBIR) is an important research area with applications to a large amount of image databases and multimedia information. In CBIR, image is described by several low-level image features, such as color, texture, shape or their combination. The focus of this paper is on the problem of shape and texture feature extraction and representation for CBIR.

Shape is one of the most important features because it is very important to human perception. Two shape representation methods, region based shape descriptors and contour based shape descriptors have been studied. A new shape descriptor, Generic Fourier Descriptor (GFD) has been discussed and evaluated with existing shape descriptors. Texture is a global feature that could be used to associate related shapes. Gabor filters (GF) has been selected for texture feature extraction. We apply Generic Fourier Descriptor (GFD) for shape feature extraction and Gabor Filters (GF) for texture feature extraction on three different databases, and we successfully combined GFD and GF together for shape and texture feature extraction. Experimental results show that the proposed GFD+GF is robust to all the test databases with best retrieval rate.

## ACKNOWLEDGEMENTS

This thesis could not be finished without the help and support of many people who are gratefully acknowledged here.

At the very first, I'm honored to express my deepest gratitude to my dedicated supervisor, Dr. J. Wu; his patient and meticulous guidance and invaluable suggestions are indispensable to the completion of this thesis.

What's more, I wish to extend my thanks to my committee members, for their support of this study. I owe special thanks to Dr. J. Chen and Dr. H. Wu, for their priceless comments on this study.

Thanks are also due to my postgraduate friends, who never failed to give me great encouragement and suggestions. Special thanks should go to Mr. Zhengqiao Ji, Ms. Lan Xu, Miss Kathleen Dizon, and Mr. Tong Wang for their help and support when I had problems writing this thesis.

At last but not least, I would like to thank my family for their support all the way from the very beginning of my postgraduate study. I am thankful to all my family members for their understanding and love.

## TABLE OF CONTENTS

ABSTRACT.....	iii
ACKNOWLEDGEMENTS.....	iv
LIST OF TABLES.....	vii
LIST OF FIGURES .....	viii

### CHAPTER

#### I. INTRODUCTION

1.1 Problem Statement and Background.....	1
1.2 Research Objectives.....	3
1.3 Thesis Outline .....	5

#### II. OVERVIEW OF CONTENT-BASED IMAGE RETRIEVAL

2.1 Introduction to content-based image retrieval.....	7
2.2 Image Content Descriptors.....	8
2.2.1 <i>Color</i> .....	8
2.2.2 <i>Texture</i> .....	10
2.2.3 <i>Shape</i> .....	11
2.3 Similarity and Performance Measurement.....	12
2.3.1 <i>Similarity Measurement</i> .....	13
2.3.1.1 Minkowski-Form Distance .....	14
2.3.1.2 $\chi^2$ Statistics .....	14
2.3.2 <i>Performance Measurement</i> .....	15
2.3.2.1 Recall and Precision Pair (RPP) .....	15
2.3.2.2 Percentage of Weighted Hits (PWH) .....	15
2.4 Image Retrieval Systems.....	16
2.5 Summary .....	17

#### III. FEATURE REPRESENTATION TECHNIQUES

3.1 General Shape Representation .....	18
3.1.1 <i>Region-based Shape Descriptor</i> .....	19
3.1.2 <i>Contour-based Shape Descriptor</i> .....	21
3.2 Generic Fourier Descriptor .....	27

3.2.1 Polar Fourier transform.....	28
3.2.2 Implementation of Generic Fourier Descriptor .....	29
3.3 General Texture Representation.....	31
3.3.1 Gabor Filters .....	31
3.4 Summary .....	34
<b>IV. APPLICATIONS IN GENERAL IMAGE RETRIEVAL</b>	
4.1 Introduction .....	36
4.2 A Review of Image Segmentation .....	37
4.3 Image Retrieval Using Shape Features .....	41
4.4 Image Retrieval Using Texture Features.....	50
4.5 Propose combined Shape and Texture Features.....	52
4.6 Summary .....	53
<b>V. EVALUATION AND EXPERIMENTAL RESULTS</b>	
5.1 Comparisons on Feature Extraction Algorithms.....	54
5.2 Similarity Measurements .....	65
5.3 Summary .....	69
<b>VI. CONCLUSIONS AND FUTURE WORK</b>	
6.1 Conclusions .....	70
6.2 Future Works.....	72
<b>REFERENCES.....</b>	<b>73</b>
<b>VITA AUCTORIS .....</b>	<b>78</b>

## LIST OF TABLES

TABLE 2.1 EXAMPLES OF CBIR SYSTEMS.....	17
TABLE 5.1 RETRIEVAL RATE ON LEMS SILHOUETTE DATABASE .....	57
TABLE 5.2 RETRIEVAL RATE ON FINGERPRINT DATABASE .....	60
TABLE 5.3 RETRIEVAL RATE ON ALOI DATABASE .....	64
TABLE 5.4 RETRIEVAL RATE ON LEMS SILHOUETTE DATABASE OF DIFFERENT SIMILARITY MEASUREMENTS .....	66
TABLE 5.5 TIME TAKEN FOR THE RETRIEVAL ON LEMS SILHOUETTE DATABASE .....	66
TABLE 5.6 RETRIEVAL RATE ON ALOI DATASET OF DIFFERENT SIMILARITY MEASUREMENTS .....	68
TABLE 5.7 TIME TAKEN FOR THE RETRIEVAL ON ALOI DATASET .....	68



## LIST OF FIGURES

FIGURE1.1: AN EXAMPLE OF TEXT-BASED IMAGE SEARCH METHOD .....	2
FIGURE1.2: IMAGE DESCRIPTION USING TEXT AND CONTENT INFORMATION .....	3
FIGURE2.1: CONTENT-BASED RETRIEVAL SYSTEM STRUCTURE.....	8
FIGURE3.1 EXAMPLES OF REGION AND CONTOUR-BASED SHAPE SIMILARITY .....	18
FIGURE3.2 EXAMPLES OF SHAPES THAT CAN BE EFFICIENTLY DESCRIBED BY THE REGION-BASED DESCRIPTOR.....	19
FIGURE3.3 EXAMPLES OF SHAPES THAT CAN BE EFFICIENTLY DESCRIBED BY THE CONTOUR- BASED DESCRIPTOR .....	21
FIGURE 3.4 EXAMPLE OF COORDINATE TRANSFORMATION USING TPS .....	23
FIGURE3.5 AN EXAMPLE FOR OBTAINING BOUNDARY COORDINATES .....	24
FIGURE3.6 THREE APPLE SHAPES AND THEIR CENTROID DISTANCE SIGNATURES .....	25
FIGURE3.7 THE FOURIER RECONSTRUCTED APPLE 1, 2, AND 3 USING:(A) THE FIRST THREE FD (B) THE FIRST FIVE FD .....	26
FIGURE3.8 BLOCK DIAGRAM OF COMPUTING FD .....	27
FIGURE3.9 (A) A SHAPE WITHOUT CONTOUR (B), (C) TWO SHAPES WITH SAME CONTOUR BUT WITH DIFFERENT INTERIOR CONTENT .....	28
FIGURE3.10 POLAR REPRESENTATION OF AN IMAGE .....	29
FIGURE3.11 FWHM SAMPLING OF SPECTRAL RESPONSES OF GABOR FILTERS IN FREQUENCY PLANE .....	33
FIGURE4.1 (A) ROBERT OPERATOR (B) SOBEL OPERATOR .....	38
FIGURE4.2 AN EXAMPLE FOR DETECTING EDGES BY CANNY EDGE DETECTOR .....	39

FIGURE4.3 QUADTREE DECOMPOSITION.....	41
FIGURE4.4 (A) ORIGINAL IMAGE (B) IMAGE SEGMENTATION USING WATERSHED	
ALGORITHM.....	41
FIGURE4.5 30 QUERY IMAGES.....	43
FIGURE4.6 EXAMPLE RETRIEVALS USING FD ON DIFFERENT NUMBER (A1, A2) 5 FDs	
(B1, B2) 10FDs (C1, C2)15FDs.....	45
FIGURE4.7 COMPARISON OF RETRIEVAL EFFECTIVENESS OF 5 FDs, 10 FDs AND 15 FDs .	46
FIGURE4.8 EXAMPLE RETRIEVALS USING GFD AND FD (A1, A2) 36GFDs (B1, B2) 60GFDs	
(C1, C2) 10FDs.....	48
FIGURE 4.9 COMPARISON OF RETRIEVAL EFFECTIVENESS OF 10FDs, 36GFDs	
AND 60GFDs .....	49
FIGURE4.10 EXAMPLE RETRIEVALS USING GF .....	51
FIGURE4.11 RETRIEVAL PERFORMANCE USING GABOR FILTERS .....	51
FIGURE5.1 EXAMPLE RETRIEVALS USING COMBINED FEATURES WITH GFD AND FD	
(A) 60GFDs (B) 36GFDs (C) GF (D) 60GFDs+GF (E) 36GFDs+GF .....	56
FIGURE5.2 COMPARISON OF RETRIEVAL EFFECTIVENESS OF GF, 36GFDs,	
60GFDs, 36GFDs+GF, 60GFDs+GF .....	56
FIGURE5.3 EXAMPLE RETRIEVALS USING COMBINED FEATURES WITH GFD AND FD	
(A) 60GFDs (B) 36GFDs (C) GF (D) 60GFDs+GF (E)36GFDs+GF.....	59
FIGURE5.4 COMPARISON OF RETRIEVAL EFFECTIVENESS OF GF, 36GFDs,	
60GFDs, 36GFDs+GF,60GFDs+GF.....	60
FIGURE5.5 ALOI IMAGE DATABASE SAMPLES .....	61

FIGURE5.6 EXAMPLE RETRIEVALS USING COMBINED FEATURES WITH GFD AND FD	
(A) 60GFDs (B) 36GFDs (C) GF (D) 60GFDs+GF (E) 36GFDs+GF .....	63
FIGURE5.7 COMPARISON OF RETRIEVAL EFFECTIVENESS OF GF, 36GFDs, 60GFDs,	
36GFDs+GF, 60GFDs+GF.....	63
FIGURE5.8 RETRIEVAL PERFORMANCE OF DIFFERENT SIMILARITY MEASUREMENTS ON	
LEMS DATASET.....	66
FIGURE5.9 RETRIEVAL PERFORMANCE OF DIFFERENT SIMILARITY MEASUREMENTS ON	
ALOI DATASET .....	67

## CHAPTER I

### INTRODUCTION

#### 1.1. Problem Statement and Background

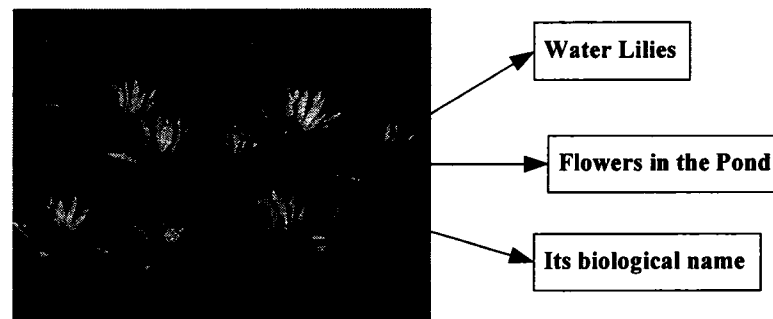
Recent years have seen the use of images in human communication increased dramatically. Researchers use related images for investigation, reporters use images for demonstration and improving text information, doctors use images to diagnose and distinguish diseases, students use images for entertaining and studying and so on.

Various image applications lead to the generation and prosperity of digital images. Both military and civilian equipment generate giga-bytes of images everyday, however, we cannot access or make use of the image information unless it is organized to allow efficient browsing, searching, and retrieval, so there has been urgent demand for techniques to index and manage digital images.

Before the 1980s, images were stored in text-based databases. Text-based image indexing runs faster and offers more flexibility because the computer does not expend resources on processing the images. For the same reason, text-based applications use memory more efficiently. Textual indexing of images provides user with key word searching, catalogue browsing and even with query interface such as Structural Query Language (SQL). A very popular framework of image retrieval was to first annotate the images by text and then use text-based database management systems (DBMS) to perform image retrieval. Two comprehensive surveys on this topic are [1, 2]. Here is a traditional method for searching an image using text-based method.

Figure 1.1 shows a common way for finding images by text-based method. In order to obtain the required left image in the figure, user can provide key words such as

water lilies, flowers in the pond or the biological name etc.



**Figure 1.1: An example of text-based image search method**

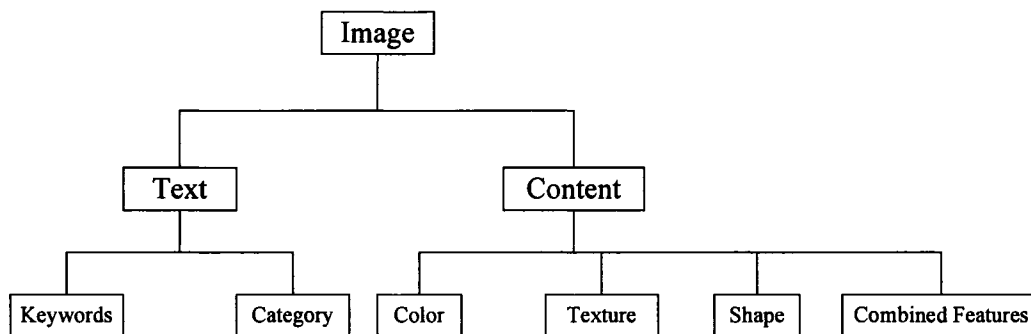
Although text-based method has its own merits, two major difficulties also exist, especially when the size of image collections is large (tens or hundreds of thousands). One is that it is time consuming in manual image annotation, when the database is large, the other one, which is more essential, is that images have rich content and are subjective to human perception. Visual features of image are difficult to be described using words, and for the same image content different people may perceive it differently. Due to the deficiencies in text-based method, in the early 1990s, Content-based Image Retrieval (CBIR) was proposed. It has the capability to retrieve image based on their content, domain independence; and can be automated. Since then, many techniques in this research direction have been developed and many image retrieval systems, both research and commercial, have been built.

Generally, in Content-based Image Retrieval, images would be indexed by their own visual content, such as color, texture and shape. One of the key issues of CBIR is the need to extract useful information from the raw data to represent the image content. So the extraction of effective content features is crucial to the success of CBIR. From the literature review, primitive features in images such as color, texture, shape or their

combinations are very useful for image description and retrieval. Feature extraction in CBIR is the most important step for successful image retrieval [3][4].

## 1.2 Research Objectives

From the above background information, we can conclude that images in databases can be represented and indexed either by textual or content information. And image content is described by primitive image features like color, texture, shape or the combination of them (Figure 1.2). This research adopts CBIR approach, representing and retrieving images by their content.



**Figure 1.2: Image Description using Text and Content information**

A successful design of a new CBIR system needs a comprehensive and good knowledge of the CBIR techniques. A basic CBIR system is mainly composed of three components: Image Pre-processing, Feature Extraction and Query Engine. Shape is one of the most important low level image features due to the fact that shape is a very important feature to human perception. Human beings tend to perceive scenes as being composed of individual objects, which can be best identified by their shapes. In our research, we first focus on shape-based image retrieval techniques, from the most classical techniques to the latest and effective techniques. The most important part in shape based image retrieval is shape feature extraction. The purpose of the feature

extraction is to have a simple and reliable representation of the shape image by eliminating the redundant information and retaining most of important clues for image retrieval. Shape descriptors are generally used for shape feature extraction. Region-based Shape Descriptor and Contour-based Shape Descriptor are two general categories in shape image feature extraction. Fourier Descriptor (FD) and Generic Fourier Descriptor (GFD) will be implemented and compared in different image databases. The extracted features are represented as shape descriptors. Shape descriptors are some set of numbers that are produced to describe a given shape. So it is easily discriminated by similarity measures. Query engine is used to calculate distance or similarity between images. The smaller the distance is, the more similar between images. One of the most popular similarity measures is Euclidean Distance. City block distance and  $\chi^2$  Statistics will be also discussed and implemented for comparison. At last, we will adopt recall and precision pair (RPP) performance measure to evaluate different algorithms for image retrieval.

The main aim of this research is to develop a better methodology for feature extraction to obtain as much image information as possible, thus retrieving more relevant images from database to the query user supplied. In an image retrieval system, the more the related images retrieved, the more useful and effective the system is. As discussed before, image content can be described by color, shape and texture. Texture is also an important feature of object. Texture is a global feature that could be used to associate related shapes. When the shape of an object changes largely, its texture usually remains little change. Therefore, if shape of object is complex, object's texture information can be exploited to describe object together with object's shape feature. So it is possible to

improve system efficiency by combining texture and shape information together to extract image features. Two-dimensional Gabor filters (GF) is proved to be very effective texture feature extraction methods in literature. Ma and Manjunath evaluated the texture image annotation by various wavelet transform representations [5]. They found that GF had the best performance among the tested candidates. So we choose GF with specified orientations and frequencies to obtain texture features for image retrieval.

Our research goal is to combine shape and texture information together for feature extraction in our CBIR system. An in-depth comparison and analysis will be carried out in different databases to test the effectiveness and robustness of the proposed method. Our image databases include silhouette database collected by The Laboratory for Engineering Man/Machine Systems (LEMS) [6], Fingerprint database [7] and Amsterdam Library of Object Images (ALOI) [8]. We transfer all images in databases into grey-scale images since we ignore color information.

### 1.3 Thesis Outline

Chapter 1 presents an introduction to the background of digital image collection and the problem of traditional text-based method to index and manage digital images, and the reason and benefit of proposed CBIR. It also provides an explanation of the objectives of this research and where the motivation came from.

Chapter 2 gives a brief review of CBIR system, and provide a brief description of feature extraction and matching for color, texture and shape information. Similarity measurements and performance measurements are described and discussed. Different similarity measurements have been evaluated. Desirable similarity and



performance measurements have been selected for use in later chapters. Some famous image retrieval system has been introduced as well.

Chapter 3 focuses on the details of the feature extraction algorithms used in this research. We first give a description for the shape descriptor and its two categories: region-based and contour-based shape descriptor. Then we will concentrate on a new proposed shape descriptor, Generic Fourier Descriptor (GFD). Gabor filters for texture feature extraction will be described at last.

Chapter 4 first briefly introduce image segmentation methodology, which is an important part for simulation in real world objects. And then we give the application for shape features as well as texture features. A combined shape and texture features are proposed at last.

Chapter 5 describes the models of the sets of experiments. We will test and compare the performance of the image retrieval rate in three image databases through different feature extraction algorithms and similarity measurements. The experimental results will be discussed and the methodology for the comparison and testing will be explained.

Chapter 6 concludes the thesis and makes recommendations for future work.

## CHAPTER II

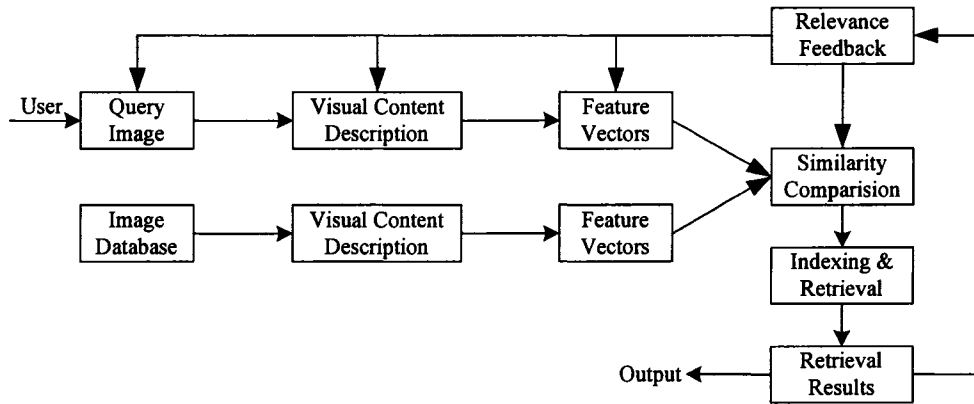
### OVERVIEW OF CONTENT-BASED IMAGE RETRIEVAL

#### 2.1. Introduction to content-based image retrieval

Content-based image retrieval (CBIR), a new technique to search images from large-scale image databases by using image contents, has been developed and prospered since the 1990s. As mentioned before, images were first annotated with text and then searched using a text-based approach from traditional database management systems. Images can be organized by tropical or semantic hierarchies to facilitate key word searching, catalogue browsing and etc. However, it is not feasible to automatically generate descriptive texts for a large-scale image databases, most text-based image retrieval system require manual annotation of images. As a result, text-based image retrieval system has its limitations; it's difficult to support a variety of task-dependent queries.

Content-based image retrieval, uses the visual contents of an image such as color, texture and shape to represent and index the image, and it can be automated and do on-line searching. Here is a typical structure of CBIR. User supplies a query image to CBIR system, the system extracts visual contents of the query images and images in the database, and then the extracted contents are described by multi-dimensional feature vectors. The feature vectors of the images in the database form a feature database while the query image is represented by its feature vectors. To retrieve images, the system calculates the similarities or distances between the feature vectors of the query image and the images in the database, and the retrieval is performed with the aid of an indexing scheme. The indexing scheme provides an efficient way to search for the image database.

Recent retrieval systems have incorporated user' relevance feedback, in order to combine human perception subjectivity into the retrieval process and provide users opportunities to evaluate retrieval results, and automatically refine queries on the basis of those evaluations. In this chapter, we introduce these fundamental techniques for content-based image retrieval.



**Figure 2.1: Content-based retrieval system structure [4]**

## 2.2 Image Content Descriptors

Generally speaking, image content may include both visual and semantic content. General visual content include color, texture, shape, etc. Semantic content is obtained either by textual annotation or by complex inference procedures based on visual content. This chapter concentrates on general visual contents descriptions, we will introduce some widely used techniques for extracting color, texture and shape information from images.

### 2.2.1 Color

The color feature is one of the most extensive used visual features in image retrieval. It is relatively robust to background complication and independent of image

size and orientation. Several methods for retrieving images on the basis of colour similarity have been described in the literature [9]. In image retrieval, the color histogram is the most commonly used color feature representation. Each image added to the collection is analysed to compute a colour histogram that shows the proportion of pixels of each colour within the image. The colour histogram for each image is then stored in the database. At search time, the user can either specify the desired proportion of each colour (75% olive green and 25% red, for example), or submit an example image from which a colour histogram is calculated. Either way, the matching process then retrieves those images whose colour histograms match those of the query most closely. The matching technique most commonly used, histogram intersection, was first developed by Swain and Ballard [10]. Variants of this technique are now used in a high proportion of current CBIR systems. Methods of improving on Swain and Ballard's original technique include the use of cumulative colour histograms [11], combining histogram intersection with some element of spatial matching [12], and the use of region-based colour querying [13]. The results from some of these systems can look quite impressive.

Besides the color histogram, several other color feature representations have been applied in image retrieval, including color moments and color sets. To overcome the quantization effects, as in the color histogram, Stricker and Orengo proposed using the color moments approach [14]. The mathematical foundation of this approach is that any color distribution can be characterized by its moments. Furthermore, since most of the information is concentrated on the low-order moments, only the first moment (mean), and the second and third central moments (variance and skewness) were extracted as the color

feature representation. Weighted Euclidean distance was used to calculate the color similarity.

### 2.2.2 Texture

Texture is another important property of images. Basically, texture representation methods can be classified into two categories: structural and statistical.

Structural methods, including morphological operator and adjacency graph, describe texture by identifying structural primitives and their placement rules. Statistical methods, including Fourier power spectra, co-occurrence matrices, shift-invariant principal component analysis (SPCA), Tamura feature, Wold decomposition, Markov random field, fractal model, and multi-resolution filtering techniques such as Gabor and wavelet transform, characterize texture by the statistical distribution of the image intensity.

Co-occurrence matrix representation of texture features was proposed by Haralick *et al* in the early 1970s [15]. This approach explored the gray level spatial dependence of texture. It first constructed a co-occurrence matrix based on the orientation and distance between image pixels and then extracted meaningful statistics from the matrix as the texture representation. After that, Tamura *et al* explored the texture representation from a different angle [16]. They developed computational approximations to the visual texture properties found to be important in psychology studies. The six visual texture properties were coarseness, contrast, directionality, linelikeness, regularity, and roughness. One major distinction between the Tamura texture representation and the co-occurrence matrix representation is that all the texture properties in Tamura

representation are visually meaningful, whereas some of the texture properties used in co-occurrence matrix representation may not be.

In the early 1990s, the wavelet transform was introduced and its theoretical framework was established, many researchers began to utilize wavelet transform for texture feature extraction. Smith and Chang used the statistics (mean and variance) extracted from the wavelet subbands as the texture representation [17, 18]. In 1996, Ma and Manjunath evaluated the texture image annotation by various wavelet transform representations [5]. They found that Gabor transform had the best performance among the tested candidates.

### 2.2.3 Shape

Shape is one of the most important low level image features due to that shape is a very important feature to human perception. Shape refers to a specific region of an image while color and texture are both global attributes of an image. A good shape representation feature for an object should be size, rotational and translation invariant in image retrieval application. Generally, shape description can be categorized into either boundary-based or region-based methods. The former expresses shape properties of the object outline (contour), the latter expresses pixel distribution within the entire object region. The most successful representatives for these two categories are Fourier Descriptor and Zernike Moments.

The main idea of Fourier descriptor is to use the Fourier transformed boundary as the shape feature, many works can be found in [19, 20], FD has many advantages over many other shape descriptors. It is simple to compute; and each descriptor has specific physical meaning; it is also simple to do normalization, making

shape matching a simple task; and it captures both global and local features. With sufficient features for selection, FD overcomes the weak discrimination ability of those simple global descriptors. FD also overcomes the noise sensitivity and difficult normalization in the shape signature representations.

Recently, several researchers have proposed the use of WD for shape description [21]. Although WD has the advantage over FD in that it is of multi-resolution in both spatial space and spectral space, the increase of spatial resolution will certainly sacrifice frequency resolution. For example, only wavelet coefficients of the few low frequencies are used to represent shape. Most importantly, the complicated matching scheme of wavelet representation makes it impractical for online shape retrieval.

CSSD has also been widely used in shape representation [22]. CSS descriptors are essentially the descriptors of key local shape features. By putting shape boundary into scale space, not only the locations of convex (or concave) segments, but also the degree of convexity (or concavity) of the segments on the shape boundary are detected.

Zernike moment, the most popular and efficient moment method in region-based shape descriptor, is to use region-based moments which are invariant to transformations as the shape feature [23]. Recently, several researchers also use grid method to describe shape [24]. The grid-based method gains lots of interest for its simple format for representation and conformance to intuition.

### 2.3 Similarity and Performance Measurement

Content-based image retrieval calculates visual similarities between a query image and images in a database according to the extracted features. The image feature

extracted is usually an N-dimensional feature vector that can be regarded as a point in a N-dimensional space. Once images are represented and indexed into the database using the extracted feature vectors, the retrieval of images is essentially the determination of similarity between the query image and the images in database, which is essentially the determination of distance between the feature vectors representing the images. Thus the retrieval results are a list of images ranked by their similarities with the query image. Many similarity measures have been developed for image retrieval based on empirical estimates of the distribution of features in recent years, the desirable distance measure should reflect human perception. So perceptually similar images should have smaller distance between them and perceptually different images should have larger distance between them. Therefore, for a given shape feature, the higher the retrieval accuracy, the better the distance measure. We will introduce several similarity measures in this section.

Another important area is performance measurement to evaluate the retrieval rate in CBIR. Generally, performance measurement measures accuracy of the retrieval. There are different performance measurements. Some of the important performance measurements will be discussed in this section and a desirable one will be chosen for our research.

### 2.3.1 Similarity Measurements

We will introduce some commonly used similarity measures. In this section, we denote  $D(I, J)$  as the distance between the query image  $I$  and the image  $J$  in the database [10][25][26].



#### 2.3.1.1 Minkowski-Form Distance

The Minkowski-Form Distance is defined as:

$$D_p(I, J) = \left( \sum_{i=0}^{N-1} |f_i(I) - f_i(J)|^p \right)^{1/p} \quad (2.1)$$

When  $p = 1$ , it is known as city block distance or Manhattan distance.

$$D_1(I, J) = \sum_{i=0}^{N-1} |f_i(I) - f_i(J)| \quad (2.2)$$

When  $p = 2$ , it is the Euclidean distance or Euclidean metric. Euclidean distance is the most widely used measure for image retrieval. The Euclidean distance is defined as follows:

$$D_2(I, J) = \left( \sum_{i=0}^{N-1} |f_i(I) - f_i(J)|^2 \right)^{1/2} \quad (2.3)$$

#### 2.3.1.2 $\chi^2$ Statistics

The  $\chi^2$  Statistics is defined as

$$D_{\chi^2}(I, J) = \sum_{i=0}^{N-1} \frac{(f_i(I) - f_i(I, J))^2}{f_i(I, J)} \quad (2.4)$$

Where  $f_i(I, J) = \frac{f_i(I) + f_i(J)}{2}$ , This quantity measures how unlikely it is that

one distribution was drawn from the population represented by the other.

### 2.3.2 Performance Measurements

To evaluate different algorithms for image retrieval, an effective performance measurement is necessary. In this section, we will introduce several different performance measurements [27][28].

#### 2.3.2.1 Recall and Precision Pair (RPP)

One of most efficient evaluation methods for image retrieval is Recall and Precision Pair (RPP), which is a standard evaluation method and popularly used by CBIR community because RPP present abundant information of retrieval result and can be easily understood. Precision measures the retrieval accuracy while recall measures the ability of retrieving relevant items from the database. In the RPP graph,

Precision is the percentage of similar images retrieved with respect to the total number of images retrieved.

$$P = \frac{\text{No. of similar images retrieved}}{\text{Total No. of images retrieved}} \quad (2.5)$$

Recall is the percentage of similar images retrieved with respect to the total number of similar images in database.

$$R = \frac{\text{No. of similar images retrieved}}{\text{Total No. of similar images in database}} \quad (2.6)$$

#### 2.3.2.2 Percentage of Weighted Hits (PWH)

PWH is similar to the recall measurement in RPP. The subjective test is the same as in RPP, that is, each subject select items relevant to the query from the dataset. However, instead of measuring recall based on binary relevance value as in RPP, PWH assigns a weighted relevance value  $w_i$  to each item in the dataset. The weight  $w_i$  is equal

to the number of subjects selecting item  $i$  as relevant or similar to the query. Therefore, PWH is defined as

$$PWH = \frac{\sum_{i=1}^n w_i}{\sum_{i=1}^N w_i} \quad (2.7)$$

Where  $n$  is the number of items returned and  $N$  is the total number of items in the database. It is easy to observe that recall is a special case of PWH when  $w_i$  takes the value of 0 and 1.

#### 2.4 Image Retrieval System

Content-based image retrieval has become a very prosperous research area since the emergence of large-scale images. Many commercial image retrieval systems have been built. Most systems allow searching the visual database contents in several different ways described as follows [4]:

- ① Interactive Browsing
- ② Search by example
- ③ Search by sketch
- ④ Search by a key word or complete text
- ⑤ Search by specification of visual features
- ⑥ Navigation with customized image categories

Table 2.1 shows a list of CBIR systems in alphabetic order, more details can be find in the corresponding website.

System	Website/ online demo
AltaVista Photofinder	<a href="http://www.altavista.com/">http://www.altavista.com/</a>
Blobworld	<a href="http://elib.cs.berkeley.edu/blobworld/start.html">http://elib.cs.berkeley.edu/blobworld/start.html</a>
CANDID	<a href="http://public.lanl.gov/kelly/CANDID/index.shtml">http://public.lanl.gov/kelly/CANDID/index.shtml</a>
MARS	<a href="http://www-db.ics.uci.edu/pages/demos/index.shtml">http://www-db.ics.uci.edu/pages/demos/index.shtml</a>
QBIC	<a href="http://wwwqbic.almaden.ibm.com/">http://wwwqbic.almaden.ibm.com/</a>
Virage	<a href="http://www.virage.com">http://www.virage.com</a> .
VisualSEEK	<a href="http://www.ctr.columbia.edu/VisualSEEK/">http://www.ctr.columbia.edu/VisualSEEK/</a>
WebSEEK	<a href="http://www.ctr.columbia.edu/webseek/">http://www.ctr.columbia.edu/webseek/</a>
WISE	<a href="http://bergman.stanford.edu/~zwang/project/imsearch/WBIIS.html">http://bergman.stanford.edu/~zwang/project/imsearch/WBIIS.html</a>

**Table2.1 Examples of CBIR systems**

## 2.5 Summary

In this chapter, we introduce some fundamental techniques for content-based image retrieval. We first give a brief introduction to CBIR, and then discuss the important parts in CBIR individually, including visual content description, similarity and performance measurement. We describe former and current techniques in each section, and introduce several algorithms as well; some of them will be discussed in next chapter for research purpose. At last, we give a brief introduction for some commercial and research image retrieval systems that are active in current CBIR area.

## CHAPTER III

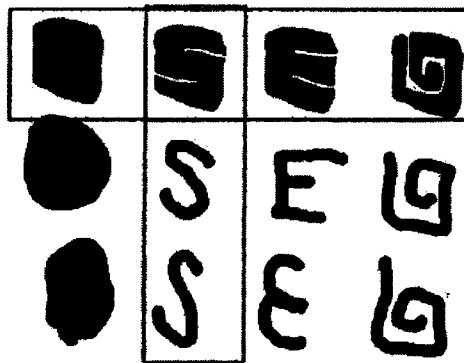
### FEATURE REPRESENTATION TECHNIQUES

#### 3.1. General Shape Representation

Shape is one of the most important low level image features due to that shape is a very important feature to human perception. Human beings tend to perceive scenes as being composed of individual objects, which can be best identified by their shapes. Besides, as far as query is concerned, shape is simple for user to describe, either by giving example or by sketching.

Shape representation generally looks for effective and perceptually important shape features based on either shape boundary information or boundary plus interior content. Various shape representation methods, or shape descriptors, exist in the literature, these methods can be classified into two categories: region based versus contour based.

The difference between region based and contour based shape descriptors can be clearly discriminated in Figure 3.1.



**Figure 3.1 Examples of Region and Contour-based shape similarity [29]**

Objects in the first row have similar spatial distribution of pixels and are therefore similar according to a region-based criterion. However, they clearly have

different outline contours. Meanwhile, objects shown in each column are similar according to a contour-based criterion. Posing a query with the object located in the first row and second column will result in retrieved object from the first row (if region-based similarity is concerned) or second column (if contour-based similarity is concerned). Now we will introduce region-based and contour-based shape descriptors for Shape-based image retrieval.

### 3.1.1 Region-based Shape Descriptor

The region-based shape descriptor expresses pixel distribution within a 2-D object region; it can describe complex objects consisting of multiple disconnected regions as well as simple objects with or without holes. The following Figure 3.2 is some object, which can be efficiently described by Region-based Shape Descriptor.



**Figure 3.2 Examples of shapes that can be efficiently described by the region-based descriptor [29]**

Region-based Shape Descriptor has some important features [29]:

- (1) It gives a compact and efficient way of describing properties of multiple disjoint regions simultaneously.
- (2) Sometimes during the process of segmentation, an object may be split into disconnected sub-regions. Such an object can still be retrieved, provided that the information which regions it was split into is retained and used during the descriptor extraction.

(3) The descriptor is robust to segmentation noise, e.g., salt and pepper noise.

Common region based methods use moment descriptors to describe shapes. These include geometric moments, Legendre moments, Zernike moments and pseudo-Zernike moments. It has been shown in [23] that Zernike moments outperforms other moment methods in terms of overall performance. Other region-based methods include grid method, shape matrix, convex hull and media axis. Here we briefly introduce Zernike moments as follows.

Zernike-moment descriptor (ZMD) is obtained by using all the pixel information within a shape region. It does not assume shape boundary information. ZMD is one of the best shape descriptors among the existing shape descriptors. It has been tested on MPEG-7 shape databases and adopted in MPEG-7 as region-based shape descriptor.

The complex Zernike moments are derived from Zernike polynomials,

$$V_{nm}(x, y) = V_{nm}(r \cos \theta, r \sin \theta) = R_{nm}(r) \exp(jm\theta) \quad (3.1)$$

$$\text{and } R_{nm}(r) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} r^{n-2s} \quad (3.2)$$

Where  $r$  is the radius from  $(x, y)$  to the shape centroid,  $\theta$  is the angle between  $r$  and  $x$ -axis,  $n$  and  $m$  are integers and subject to  $n-|m|=\text{even}$ , where  $m \leq n$ . Zernike polynomials are a complete set of complex-valued function orthogonal over the unit disk, i.e.,  $x^2 + y^2 = 1$ . Then the complex Zernike moments of order  $n$  with repetition  $m$  are defined as

$$\begin{aligned}
A_{nm} &= \frac{n+1}{\pi} \sum_x \sum_y f(x, y) V_{nm}^*(x, y) \\
&= \frac{n+1}{\pi} \sum_r \sum_{\theta} f(r \cos \theta, r \sin \theta) R_{nm}(r) \times \exp(jm\theta), r \leq 1
\end{aligned} \tag{3.3}$$

Where  $f(x, y)$  is a binary shape function,  $V^*(x, y)$  is the complex conjugate of  $V(x, y)$ . The magnitudes of the acquired Zernike moments normalized by the mass of the shape are used as shape descriptors.

### 3.1.2 Contour-based Shape Descriptor

The contour-based descriptor expresses shape properties of the object outline (contour). Objects for which characteristic shape features are contained in the contour are described efficiently by this descriptor.



**Figure 3.3 Examples of shapes that can be efficiently described by the contour-based descriptor [29]**

Contour-based Shape Descriptor has the following important features:

- (1) It can distinguish between shapes that have similar region-shape properties but different contour-shape properties.
- (2) It supports search for shapes that are semantically similar for humans, even when significant intra-class variability exists
- (3) It is robust to significant nonrigid deformations
- (4) It is robust to distortions in the contour due to perspective transformations, which are common in the images and video.



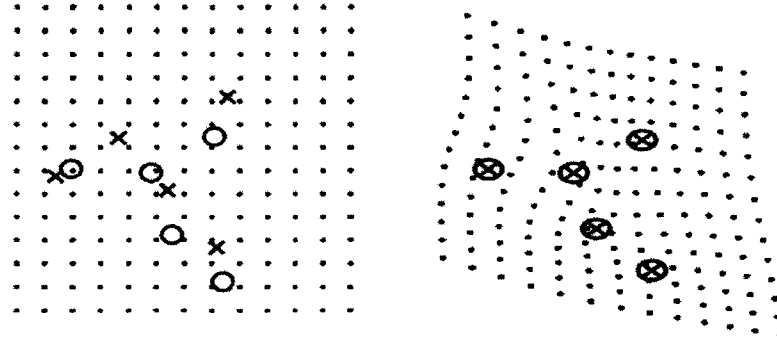
Contour shape representations include global shape descriptors such as eccentricity, circularity and bending energy, shape signatures such as chain code, centroid distance, and cumulative angles, spectral descriptors such as FD and wavelet descriptors, and curvature scale space descriptors (CSSD).

Global contour shape representation techniques usually compute a multi-dimensional numeric feature vector from the shape boundary information. Common simple global descriptors are area, circularity ( $\text{perimeter}^2/\text{area}$ ), eccentricity (length of major axis/length of minor axis), major axis orientation, and bending energy. These simple global descriptors usually can only discriminate shapes with large differences; therefore, they are usually used as filters to eliminate false hits or combined with other shape descriptors to discriminate shapes. They are not suitable to be standalone shape descriptors.

Bending energy is based on thin-plate splines to calculate the energy needed for coordinate transformation between two images. The smaller the bending energy, the more similar between two images.

The thin-plate spline (TPS) is a spatial generalization of the cubic spline for estimating the deformation that warps one set of point patterns to another in 2D space. The application of TPS to match images with deformations was first proposed by Goshtasby [30].

Figure 3.4 shows a Simple example of coordinate transformation using TPS. The bending energy between red cross points and blue circle points is the energy needed to warp the red cross points and blue circle points into a single mapping.



**Figure 3.4 Example of coordinate transformation using TPS**

A Shape signature represents a shape by a one-dimensional function derived from shape boundary points. Shape signatures are local representations of shape features and are extracted from spatial domain, as a result, they are sensitive to noise. Furthermore, the matching cost using shape signatures is high due to the complex normalization of rotation invariance. Therefore, these representations need further processing using scale space or using spectral transform such as Fourier transform, short-time Fourier transform, and wavelet transform.

Spectral descriptors overcome the problem of noise sensitivity and boundary variations by analyzing shape in spectral domain. Spectral descriptors include Fourier descriptor (FD), wavelet descriptor (WD) and curvature scale space descriptors (CSSD); they are derived from spectral transforms on 1-D shape signatures.

Many comparative researches have been done in this area. Zhang and Lu [31,32,33] have found that for general shapes, the centroid distance function is the most desirable shape signature to derive FD. They have also found that 10 FD features are sufficient to represent shape, this is a significant reduction in dimensions of FD compared with 60 FD features usually used in shape representation. Their results show that FD

outperforms CSSD method in terms of retrieval performance and robustness. Now we will discuss FD as follows.

In general, FDs are obtained by applying Fourier transform on a shape signature, the normalized Fourier transformed coefficients are called the Fourier descriptors of the shape. The shape signature is a one-dimensional function that is derived from shape boundary coordinates. Different shape signatures have been exploited to obtain FD. Complex coordinates; curvature function, cumulative angular function, and centroid distance are the commonly used shape signatures. However, FD derived from different signatures has significant different performance on shape retrieval. As has been shown in [34], FD derived from centroid distance function outperforms FD derived from other shape signatures in terms of overall performance.

The first step of computing FD is to obtain the boundary coordinates  $(x(t), y(t))$ ,  $t = 0, 1, \dots, N-1$ , where  $N$  is the number of boundary points. In order to obtain a shape boundary, we have to pre-process the image. In general, we will binarize the shape image; a simple thresholding is applied to convert the gray level shape image into binary image. And then we trace the image contour in order to obtain the boundary coordinates. Here is an example to obtain the boundary coordinates.



**Figure 3.5** An example for obtaining boundary coordinates

After obtaining the boundary coordinates, the centroid distance function can be expressed by the distance of the boundary points from the centroid  $(x_c; y_c)$  of the shape.

$$r(t) = ([x(t) - x_c]^2 + [y(t) - y_c]^2)^{1/2}, \quad t = 0, 1, \dots, N-1 \quad (3.4)$$

$$\text{Where} \quad x_c = \frac{1}{N} \sum_{t=0}^{N-1} x(t), \quad y_c = \frac{1}{N} \sum_{t=0}^{N-1} y(t) \quad (3.5)$$

For example, the following are three apple shapes and their centroid distance function  $r(t)$ , which are shown in Fig. 3.6. The left apple is the original shape, the other two are the noise affected and defected versions of the original shape. The irregular and defected shapes are likely resulted from segmentation of natural scene where noise, irregularities, and occlusions around object boundary are common.



**Fig. 3.6 Three apple shapes and their centroid distance signatures [31]**

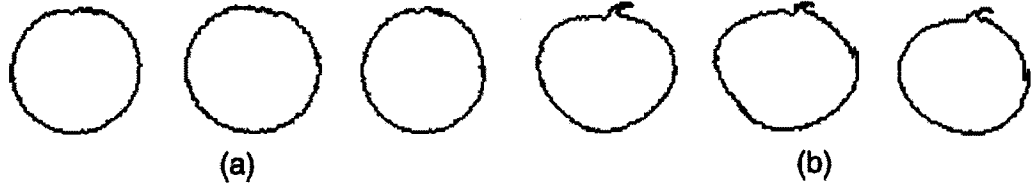
The discrete Fourier transform of  $r(t)$  is then given by

$$a_n = \frac{1}{N} \sum_{t=0}^{N-1} r(t) \exp\left(\frac{-j2\pi nt}{N}\right), \quad n = 0, 1, \dots, N-1 \quad (3.6)$$

Where  $a_n$  are the Fourier transformed coefficients of  $r(t)$ . The reconstructed shapes of apple 1, 2, and 3 using the first three coefficients are shown in Fig. 3.5a and the reconstructed shapes of apple 1, 2, and 3 using the first five coefficients are shown in

Fig. 3.7b. As can be seen, the three reconstructed shapes in Fig. 3.7a are very similar even though their original shapes are affected by serious noise and defections. The same is for the three shapes in Fig. 3.5b. Since the low frequency coefficients are

usually the most significant and are used to derive shape descriptors, the two examples indicate that, shape descriptors using Fourier transformed coefficients are robust to variations of shape boundaries.



**Fig.3.7 The Fourier reconstructed apple 1, 2, and 3 using:(a) the first three FD (b) the first five FD[31]**

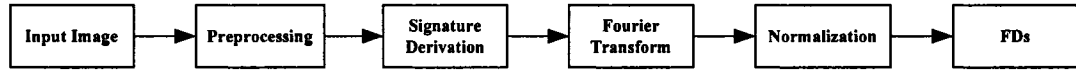
The acquired Fourier coefficients are translation invariant due to the translation invariance of the shape signature. In order to describe the shape, the acquired Fourier coefficients have to be further normalized so that they are rotation, scaling and start point independent shape descriptors. From Fourier transform theory, the general form of the Fourier coefficients of a contour  $r(t)$  transformed through translation, rotation, scaling and change of start point from the original contour  $r(t)^{(0)}$  is given by:

$$a_n = \exp(jn\tau) \cdot \exp(j\phi) \cdot s \cdot a_n^{(0)} \quad (3.7)$$

Where  $a_n$  and  $a_n^{(0)}$  are the Fourier coefficients of the transformed shape and the original shape, respectively,  $\tau$  and  $\phi$  are the angles incurred by the change of start point and the rotation, respectively;  $s$  is the scale factor. Now considering the following expression:

$$\begin{aligned} b_n &= \frac{a_n}{a_0} = \frac{\exp(jn\tau) \cdot \exp(j\phi) \cdot s \cdot a_n^{(0)}}{\exp(j\tau) \cdot \exp(j\phi) \cdot s \cdot a_0^{(0)}} \\ &= \frac{a_n^{(0)}}{a_0^{(0)}} \exp[j(n-1)\tau] = b_n^{(0)} \exp[j(n-1)\tau] \end{aligned} \quad (3.8)$$

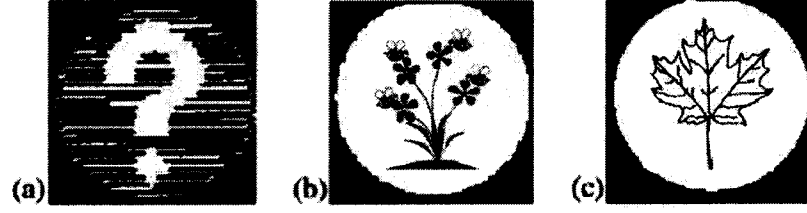
Where  $b_n$  and  $b_n^{(0)}$  are the normalized Fourier coefficients of the transformed shape and the original shape, respectively. The normalized coefficient of the derived shape  $b_n$  and that of the original shape  $b_n^{(0)}$  only have difference of  $\exp[j(n-1)\tau]$ . If we ignore the phase information and only use magnitude of the coefficients, then  $|b_n|$  and  $|b_n^{(0)}|$  are the same. In other words,  $|b_n|$  is invariant to translation, rotation, scaling and change of start point. The set of magnitudes of the normalized Fourier coefficients of the shape  $\{|b_n|, 0 < n \leq N\}$  are used as shape descriptors, denoted as  $\{FD_n, 0 < n \leq N\}$ . The whole process of computing FD from a shape is given in Fig. 3.8.



**Fig.3.8 Block diagram of computing FD [31]**

### 3.2 Generic Fourier Descriptor

We have discussed two categories of shape descriptor: region based versus contour based, and their typical descriptors: Zernike-moment descriptor (ZMD) and Fourier Descriptor (FD). However, there are several drawbacks existed in the existing descriptors. Most of the region-based shape descriptors are extracted from spatial domain, as the result, they are sensitive to noise and shape variations. And in one-dimensional FD, it is difficult to derive 1-D FD for the shape in Fig. 3.9(a), because the contour of the shape is not available. Furthermore, 1-D FD cannot capture shape interior content that is important for shape discrimination, which is show in Fig. 3.9(b) and Fig. 3.9(c); FD is not able to discriminate the shape between them.



**Fig. 3.9 (a) A shape without contour**  
**(b), (c) two shapes with same contour but with different interior content**  
 In order to overcome the disadvantages of the existing shape descriptors,

Zhang and Lu proposed a generic Fourier Descriptor (GFD) in 2002 [33]. The proposed shape descriptor is derived by applying two-dimensional Fourier transform on a polar-raster sampled shape image.

### 3.2.1 Polar Fourier transform

As has been discussed before, One-dimensional FT has been successfully applied to contour shape to derive FD. The application of one-dimensional FT on shape assumes the knowledge of shape boundary information, which may not be available in general situations. Now we introduce 2D Fourier Transform first.

Here are formulas for the continuous and discrete 2-D Fourier transform of a shape image  $f(x, y)$  ( $0 \leq x < M, 0 \leq y < N$ )

$$F(u, v) = \int_x \int_y f(x, y) \times \exp[-j2\pi(ux + vy)] dx dy \quad (3.9)$$

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \times \exp[-j2\pi(ux / M + vy / N)] \quad (3.10)$$

However, direct applying the 2-D FT to the Cartesian representation of an image is not practical. Because the resulting descriptors are not rotation invariant, which is a crucial property for a shape descriptor. Zhang et al. give a solution to this problem by applying 2-D FT on polar-raster sampled shape image. Furthermore, In order to find a

correspondence between input parameters and their physical meaning, Zhang et al. proposed to treat the polar image in polar space as a normal two-dimensional rectangular image in Cartesian space, as shown in Fig. 3.10.

If we apply 2-D FT on the rectangular image, the polar FT has the similar form to the normal 2-D discrete FT of (3.10) in Cartesian space. Consequently, the modified polar FT is obtained as:

$$PF(\rho, \varphi) = \sum_r \sum_i f(r, \theta_i) \times \exp[-j2\pi(\frac{r}{R}\rho + \frac{2\pi i}{T}\varphi)] \quad (3.11)$$

Where  $0 \leq r < R$  and  $\theta_i = i(2\pi/T)$  ( $0 \leq i < T$ );  $0 \leq \rho < R$  and  $0 \leq \varphi < T$ .  $R$  and  $T$  are the radial and angular resolutions.  $f(x, y)$  is a binary function in shape application. From (3.11), the physical meaning of  $\rho$  and  $\varphi$  are simply the number of radial frequencies selected and the number of angular frequencies selected. The determination of  $\rho$  and  $\varphi$  is physically achievable, because shape features are usually captured by the few low frequencies.



Figure 3.10 Polar representation of an image

### 3.2.2 Implementation of Generic Fourier Descriptor

The implementation of GFD can be summarized into four steps, translation normalization, polar Fourier transform, rotation normalization and scale normalization. Now we will discuss the procedures in detail.



Given a shape image  $I = \{f(x, y); 0 \leq x < M, 0 \leq y < N\}$ , the shape image is first converted from Cartesian space to polar space  $I_p = \{f(r, \theta); 0 \leq r < R, 0 \leq \theta < 2\pi\}$ .  $R$  is the maximum radius of the shape. The origin of the polar space is set to be the centroid of the shape, so that the shape is translation invariant. The centroid  $(x_c, y_c)$  is given by

$$x_c = \frac{1}{M} \sum_{x=0}^{N-1} x, \quad y_c = \frac{1}{N} \sum_{y=0}^{M-1} y \quad (3.12)$$

$$\text{and } r = ([x - x_c]^2 + [y - y_c]^2)^{1/2}, \theta = \arctan \frac{y - y_c}{x - x_c} \quad (3.13)$$

The acquired coefficients of the polar Fourier transform are translation invariant due to the use of centroid as polar space origin. Rotation invariance is achieved by ignoring the phase information in the coefficients and only retaining the magnitudes of the coefficients. To achieve scale invariance, the first magnitude value is normalized by the area of the circle (area) in which the polar image resides or the mass of the shape (mass), and all the other magnitude values are normalized by the magnitude of the first coefficient. The translation, rotation and scale normalized PFT coefficients are used as the shape descriptors.

$$FD = \left\{ \frac{|PF(0,0)|}{mass}, \frac{|PF(0,1)|}{|PF(0,0)|}, \dots, \frac{|PF(0,n)|}{|PF(0,0)|}, \dots, \frac{|PF(m,0)|}{|PF(0,0)|}, \frac{|PF(m,n)|}{|PF(0,0)|} \right\} \quad (3.14)$$

For efficient shape description, only a small number of the acquired descriptors are selected for shape representation. The selected descriptors form a feature vector that is used for indexing the shape. For two shapes represented by their Fourier descriptors, the similarity between the two shapes is measured by the similarity measurement between the two feature vectors of the shapes.

### 3.3 General Texture Representation

Texture refers to the visual patterns that have properties of homogeneity that do not result from the presence of only a single color or intensity [18]. It is an innate property of virtually all surfaces, including clouds, trees, bricks, hair, and fabric. It contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment [15].

Texture is also an important feature of object. Texture is a global feature that could be used to associate related shapes. When shape of object changes largely, its texture usually remains little changed. Therefore, if shape of object is complex, object's texture information can be exploited to describe object together with object's shape feature. So texture features could be a good supplement for shape feature extraction in image retrieval.

Previous attempts for extracting texture features include the following approaches: random field modeling [35], co-occurrence matrices [15] and Gabor filters [5]. Two-dimensional Gabor filters (GF) is proved to be very effective texture feature extraction methods in literature. Ma and Manjunath evaluated the texture image annotation by various wavelet transform representations [5], including orthogonal and bi-orthogonal wavelet transforms, the tree-structured wavelet transform, and the Gabor wavelet transform. They found that the Gabor transform was the best among the tested candidates which matched human vision study results.

#### 3.3.1 Gabor Filters

Gabor filters are a group of wavelets, with each wavelet capturing energy at a specific frequency and direction. Typically, an input image  $I(x, y)$  with size  $P \times Q$ , is

convolved with a 2D Gabor function  $g_{mn}(x, y)$ , to obtain a Gabor feature  $G_{mn}(x, y)$  as follows:

$$G_{mn}(x, y) = \sum_{x_1} \sum_{y_1} I(x_1, y_1) g_{mn}^*(x - x_1, y - y_1) \quad (3.15)$$

Where \* indicates the complex conjugate.

A 2D Gabor function  $g(x, y)$  has its form:

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi jWx\right] \quad (3.16)$$

And  $g_{mn}(x, y)$  is a set of self-similar functions generated from dilation and rotation of the Gabor function  $g(x, y)$  [5].

$$g_{mn}(x, y) = a^{-m} g(x', y') \quad (3.17)$$

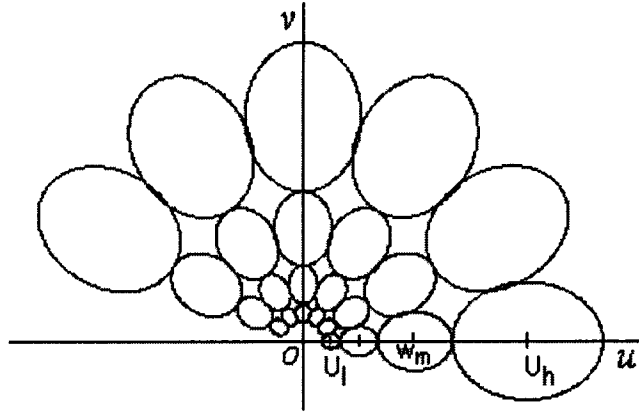
$$x' = a^{-m}(x \cos \theta + y \sin \theta), \quad y' = a^{-m}(-x \sin \theta + y \cos \theta) \quad (3.18)$$

Where m and n specify the number of scales and orientations respectively ( $m = 0, 1, \dots, M-1, n = 0, 1, \dots, N-1$ ), and  $a > 1, \theta = n\pi / N$ .

In order to decide the bank of Gabor filters, the parameters in (3.16) to (3.18) have to be determined. The Gabor wavelets generated using (3.16) through (3.18) are complete but not orthogonal wavelets, it implies there is redundancy in the filters. Therefore we may design a bank of Gabor filters so that redundancy will be significantly reduced while image texture will be well represented by the set of individual filter response. Due to the identical functional form of Gabor response function  $g(x, y)$  and its spectral function, its 2D Fourier transform  $G(u, v)$  can be written as

$$G(u, v) = \exp\left\{-\frac{1}{2}\left[\frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right]\right\} \quad (3.19)$$

Where  $\sigma_u = (2\pi\sigma_x)^{-1}$  and  $\sigma_v = (2\pi\sigma_y)^{-1}$ . An optimal representation of image in spatial domain can be achieved by finding an optimal representation of image frequency spectrum. The idea of achieving so is to use the full width at half maximum (FWHM) of the Gabor spectral functions to form a complete coverage of the frequency plane within the modulation frequency bandwidth. Specifically, the design follows three principles [26]: (i) Uniform sampling of orientation bandwidth; (ii) Exponential sampling of modulation bandwidth; (iii) Continuous coverage of the frequency space. This strategy is illustrated in Figure 3.11.



**Figure 3.11 FWHM sampling of spectral responses of Gabor filters in frequency plane**

Let  $U_l$  and  $U_h$  denote the lower and upper center frequencies of interest. Let  $K$  be the number of orientations and  $S$  be the number of scales in the multiresolution decomposition. By the above strategy of orthogonalization, the variables in the above equations can be determined as following:

$$a = (U_h / U_l)^{\frac{1}{S-1}}$$

$$\sigma_u = \frac{(a-1)U_h}{(a+1)\sqrt{2 \ln 2}}$$

$$\sigma_v = \tan\left(\frac{\pi}{2k}\right)[U_h - 2 \ln\left(\frac{\sigma_u^2}{U_h}\right)][2 \ln 2 - \frac{(2 \ln 2)^2 \sigma_u^2}{U_h^2}]^{\frac{1}{2}} \quad (3.20)$$

We can obtain a set of magnitudes by applying Gabor filters on the image  $I(x, y)$  with different orientation at different scale.

$$E(m, n) = \sum_x \sum_y |G_{mn}(x, y)| \quad (3.21)$$

The mean  $\mu_{mn}$  and standard deviation  $\sigma_{mn}$  of the magnitude of the transformed coefficients are as follows:

$$\mu_{mn} = \frac{E(m, n)}{P \times Q}, \quad \sigma_{mn} = \frac{\sqrt{\sum_x \sum_y (|G_{mn}(x, y)| - \mu_{mn})^2}}{P \times Q} \quad (3.22)$$

The Gabor feature vector is given by:

$$f = [\sigma_{00}, \sigma_{01}, \dots, \sigma_{(M-1)(N-1)}] \quad (3.23)$$

For two images represented by their Gabor feature vectors, the similarity between the two images is measured by the similarity measurement between the two feature vectors.

### 3.4 Summary

In this chapter, State of the art shape representation techniques have been reviewed. Two shape representation methods, region based shape descriptors and contour based shape descriptors have been introduced. Zernike moments and Fourier descriptor has been described in detail respectively. A Generic Fourier Descriptor, which overcomes

the drawbacks of existing shape representation techniques, has been introduced. Gabor filters for texture feature extraction has been discussed at last.

## CHAPTER IV

### APPLICATIONS IN GENERAL IMAGE RETRIEVAL

#### 4.1 Introduction

In the previous three chapters, we have described content-based image retrieval and its composition. A great amount of effort has been put on shape representation and retrieval. It is assumed that the shape object has already been available and existed in binary form. The shapes in the database are actually the simulation of real world objects and their variations under different conditions. In reality, objects more likely appear in general images with heterogenous background. The shapes of segmented objects from general image can be more complex and inaccurate than shapes from homogenous shape database. Therefore, a practical shape description and retrieval scheme should take segmented objects into consideration.

As discussed before, shape descriptors can be broadly classified in two categories: Common-based and Region-based. Common based methods use image contours to describe shape. In region-based techniques, shape descriptors are derived using all the pixel information within a shape region.

Although region-based shape representations can be applied to more general situations, they usually involve more computation. On the other hand, it has been known that contours are so dominant in visual perception and other area. Some comparative study has been done in common contour-based methods, and the result is that FD outperforms other methods in terms of robustness, low computation, retrieval performance and so on.

In this chapter, FD and GFD are applied to retrieval of objects in general images. The retrieval effectiveness of GFD and the retrieval effectiveness of FD will also be compared.

Texture is also an important feature in CBIR, Gabor texture feature described in last chapter will be applied to retrieval of objects for texture image databases.

When shape of object changes largely, its texture usually remains little changed. Intuition tells that combining object shape feature with texture feature gives more accurate description of object. Therefore, general image retrieval using combined shape features and texture features will be tested.

In order to apply shape features for general image retrieval, a practical segmentation scheme is needed to separate objects from the image. In the rest of the chapter, a brief review of current segmentation techniques is first given in Section 4.2. Image indexing using shape features and texture features are described in Section 4.3 and Section 4.4 respectively. The combined shape and texture features for image retrieval are discussed in section 4.5 at last.

#### 4.2 A Review of Image Segmentation

Generally, there are two types of image segmentation methods, contour based methods and region based methods. The two methods result in different separations of the image.

Contour based segmentation usually refers to edge detection. Edges typically occur on the boundary between two different regions in an image. Edges are found by using edge operators. Common operators are Robert operator, Sobel operator, Laplacian



operator and Canny operator. These operators are derived based on the derivatives of image intensity or on the derivatives of Gaussian [36].

$$\text{The gradient of an image: } \nabla f = \left[ \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right] \quad (4.1)$$

In order to differentiate a digital image  $f(x, y)$ , we will take discrete derivative (finite difference).

$$\frac{\partial f}{\partial x} \approx f(x+1, y) - f(x, y) \quad (4.2)$$

Robert operator and Sobel operator are based on spatial gradient that is the first derivative of image intensity. Robert operator finds edge points at gradient discontinuities in diagonal directions while Sobel operator finds edge points at gradient discontinuities in horizontal and vertical directions.

$\Delta_1$ $\begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix}$	$\Delta_2$ $\begin{matrix} 1 & 0 \\ 0 & -1 \end{matrix}$	$\Delta_1$ $\begin{matrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{matrix}$	$\Delta_2$ $\begin{matrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{matrix}$
(a)		(b)	

**Figure 4.1 (a) Robert Operator (b) Sobel Operator**

Laplacian operator, denoted by  $\Delta$  or  $\nabla^2$ , is a second order differential operator. It is described as follows:

$$\nabla^2 I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad (4.3)$$

Laplacian operator is based on finding zero-crossings of the second derivative of image intensity. Since edge points found based on the second derivatives are very sensitive to noise, it is desirable to filter out the noise before edge enhancement. Marr and

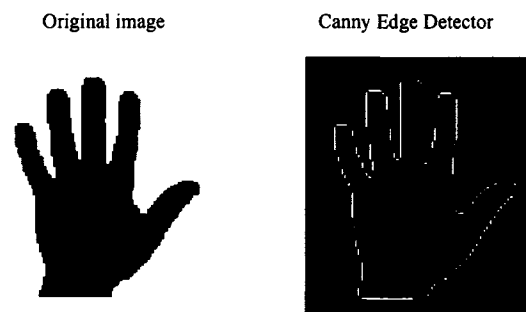
Hildreth combines Gaussian filtering with Laplacian for edge detection [37].

The Canny edge detector is based on the first derivative of the Gaussian [38]. It was designed to be an optimal edge detector; closely approximates the operator that optimizes the product of signal-to-noise ratio and localization.

The Canny edge detector works in a multi-stage process. First of all, the image is smoothed by Gaussian convolution. Then a simple 2-D first derivative operator is applied to the smoothed image to highlight regions of the image with high first spatial derivatives. Edges give rise to ridges in the gradient magnitude image. This stage is known as *canny-enhancer*.

The algorithm then tracks along the top of these ridges and sets to zero all pixels that are not actually on the ridge top so as to give a thin line in the output, a process known as *non-maximal suppression*.

The tracking process exhibits *hysteresis thresh* controlled by two thresholds:  $T1$  and  $T2$ , with  $T1 > T2$ . Tracking can only begin at a point on a ridge higher than  $T1$ . Tracking then continues in both directions out from that point until the height of the ridge falls below  $T2$ . This hysteresis helps to ensure that noisy edges are not broken up into multiple edge fragments.



**Figure 4.2 An Example for detecting edges by Canny Edge Detector**

Region-based Segmentation is to divide an image into zones of maximum homogeneity. The main idea here is to classify a particular image into a number of regions or classes. In general, it can be expressed as [39]

$$R = \bigcup_{i=1}^N R_i \quad R_i \cap R_j = \emptyset, \quad i \neq j \quad (4.4)$$

$$H(R_i) = TRUE \quad i = 1, 2, \dots, N \quad (4.5)$$

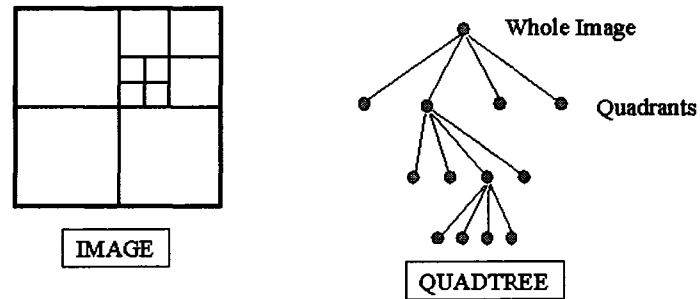
$$H(R_i \cup R_j) = FALSE, i \neq j, R_i \text{ adjacent to } R_j \quad (4.6)$$

Where  $N$  is the total number of regions in an image and  $H(R_i)$  is a binary homogeneity evaluation of the region  $R_i$ . Resulting regions of the segmented image must be both homogeneous and maximal. There are several region based segmentation methods; they include region growing, split and merge, clustering, and higher level grouping.

Region Growing is an approach to image segmentation in which neighbouring pixels are examined and added to a region class if no edges are detected. Region growing is then a process that divides image pixels into a set of adjacent regions whose feature variation within the region does not exceed a given threshold [40]. Recently, morphological segmentation using watershed algorithm has attracted much attention [41]. A typical morphological segmentation technique consists of three main steps: image simplification, marker extraction and region growing using watershed algorithm.

Split and merge technique is a top-down approach in contrast to the bottom-up approach in region growing. An important data structure which is used in split and

merge algorithms is the quadtree [42]. Figure 4.3 shows a quadtree and its relation to the image.



**Figure 4.3 Quadtree Decomposition**

Clustering is based on optimization algorithm. The widely known clustering techniques are k-Means clustering and Bayesian clustering. K-Means clustering is a segmentation method based on the minimization of the sum of squared distances from all points in a cluster to a cluster center [42]. Bayesian clustering is a segmentation method based on statistical modeling of image regions.



**Figure 4.4 (a) Original image (b) Image segmentation using watershed algorithm**

### 4.3 Image Retrieval Using Shape Features

The previous section describes how objects are segmented from an image and then indexed to database. For each object segmented, a shape region is obtained and the object is converted to binary image. The shape can be indexed using either contour-based shape descriptor or region-based shape descriptor.

Although region-based shape representations can be applied to more general situations, they usually involve more computation. On the other hand, it has been known that contours are so dominant in visual perception and other area. Some comparative study has been done in common contour-based methods, and the result is that FD outperforms other methods in terms of robustness, low computation, retrieval performance and so on.

Therefore, each image is indexed to database using FD features. To compare with FD on shape image retrieval, each image is also indexed with GFD features.

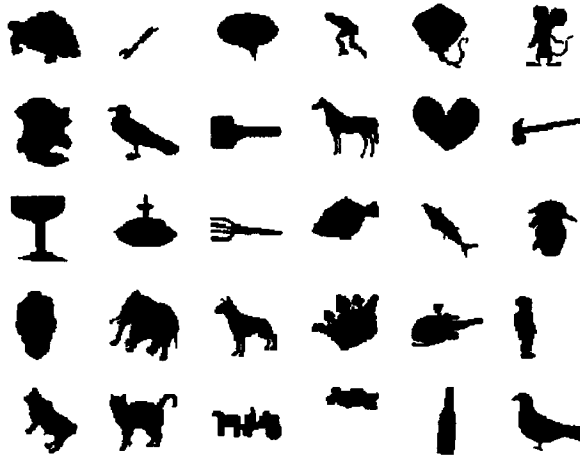
To test the retrieval performance of the FD and GFD, the retrieval test is conducted on a contour shape database collected by The Laboratory for Engineering Man/Machine Systems (LEMS) in Brown University [6]. Our silhouette database consists of shapes acquired from real world objects. It is composed of 600 shapes that have been classified to 30 classes.

Another important factor is parameter selection for FD and GFD. As discussed in section 3.1.2, shape features extracted by FD are a set of Fourier coefficients  $\{FD_n, 0 < n \leq N\}$ ,  $N$  is the number of FDs needed for shape description. Based on literature results [31], we test retrieval performance using different number of FDs, 5, 10 and 15 FDs are selected for shape description.

GFD also need to select appropriate parameters for image retrieval. The shape descriptor extracted by GFD has been represented in (3.14); therefore  $m$  angular frequencies and  $n$  radial frequencies are under decided. In our implementation, we use 36 GFDs (three radial frequencies and twelve angular frequencies) and 60 GFDs (four radial

frequencies and fifteen angular frequencies) indicated in literature [33] for feature extraction.

Each image in the database is indexed using FD and GFD, 30 images (one image per class) are randomly selected as queries. The 30 query images are shown in Figure 4.5.



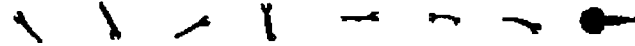
**Figure 4.5. 30 query images**

The example retrievals use FD with different number (5, 10 and 15) of FDs is showed in figure 4.6, the similarity between two FD features is measured by the Euclidean distance discussed in section 2.3.1. The top left image is the query image in our experiment.

2 (0).pgm2 (8).pgm2 (6).pgm2 (2).pgm2 (10).pgm2 (14).pgm2 (9).pgm2 (4).pgm



2 (16).pgm2 (12).pgm2 (17).pgm2 (1).pgm2 (7).pgm2 (3).pgm<sup>2</sup> (13).pgm<sub>9</sub> (4).pgm



9 (12).pgm9 (2).pgm9 (8).pgm9 (18).pgm9 (14).pgm9 (7).pgm9 (3).pgm9 (11).pgm

(a1)



2 (0).pgm2 (8).pgm2 (6).pgm2 (2).pgm2 (4).pgm2 (10).pgm2 (14).pgm2 (9).pgm



2 (1).pgm2 (16).pgm2 (12).pgm2 (7).pgm2 (17).pgm<sub>2</sub> (3).pgm<sup>2</sup> (5).pgm2 (13).pgm



9 (12).pgm<sup>2</sup> (11).pgm<sub>9</sub> (8).pgm<sup>2</sup> (18).pgm<sub>9</sub> (4).pgm9 (18).pgm9 (2).pgm9 (3).pgm

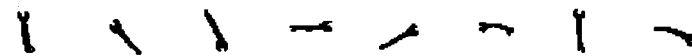
(b1)



2 (0).pgm2 (8).pgm2 (2).pgm2 (6).pgm2 (4).pgm2 (10).pgm2 (14).pgm2 (9).pgm



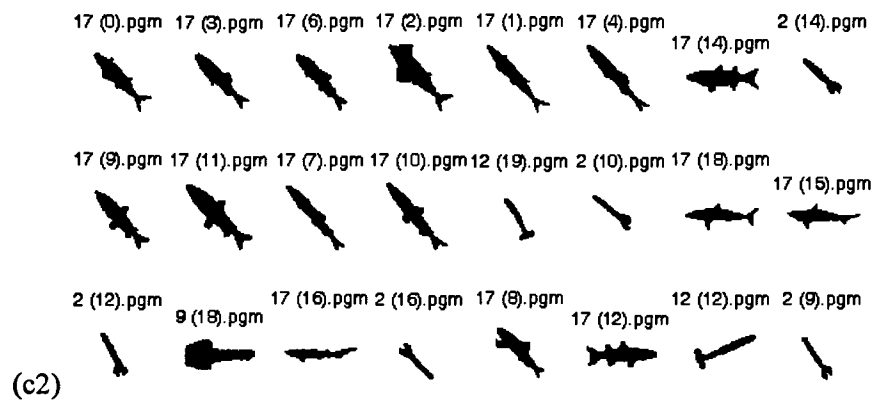
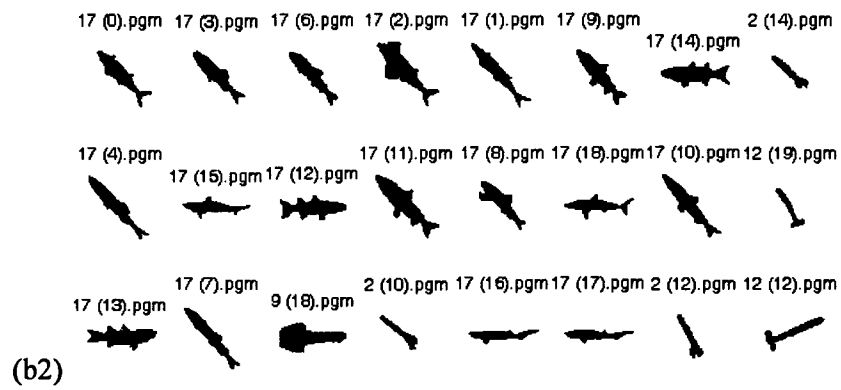
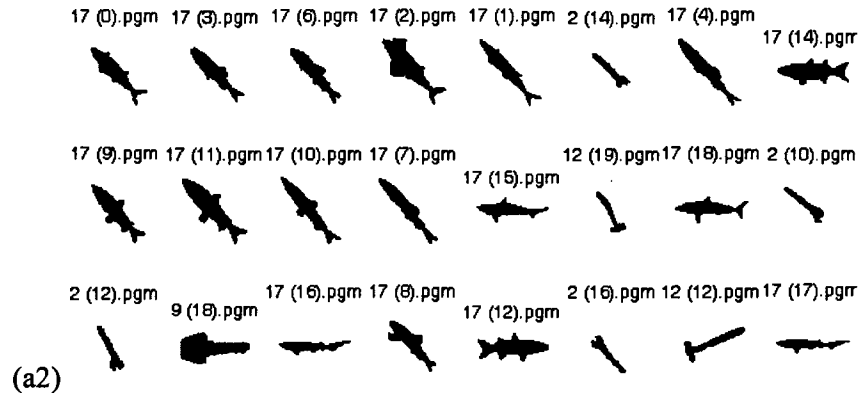
2 (1).pgm2 (16).pgm2 (12).pgm2 (7).pgm2 (17).pgm<sub>2</sub> (3).pgm<sup>2</sup> (5).pgm2 (13).pgm



9 (12).pgm<sup>2</sup> (11).pgm<sub>9</sub> (8).pgm<sup>2</sup> (18).pgm<sub>9</sub> (4).pgm9 (18).pgm9 (2).pgm9 (3).pgm

(c1)



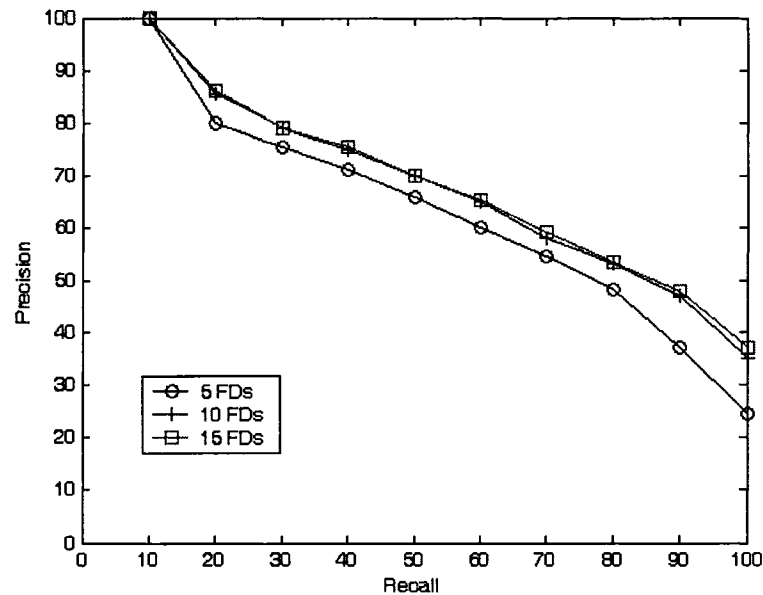


**Figure 4.6 Example retrievals using FD on different number**

**(a1, a2) 5 FDs (b1, b2) 10FDs (c1, c2) 15FDs**



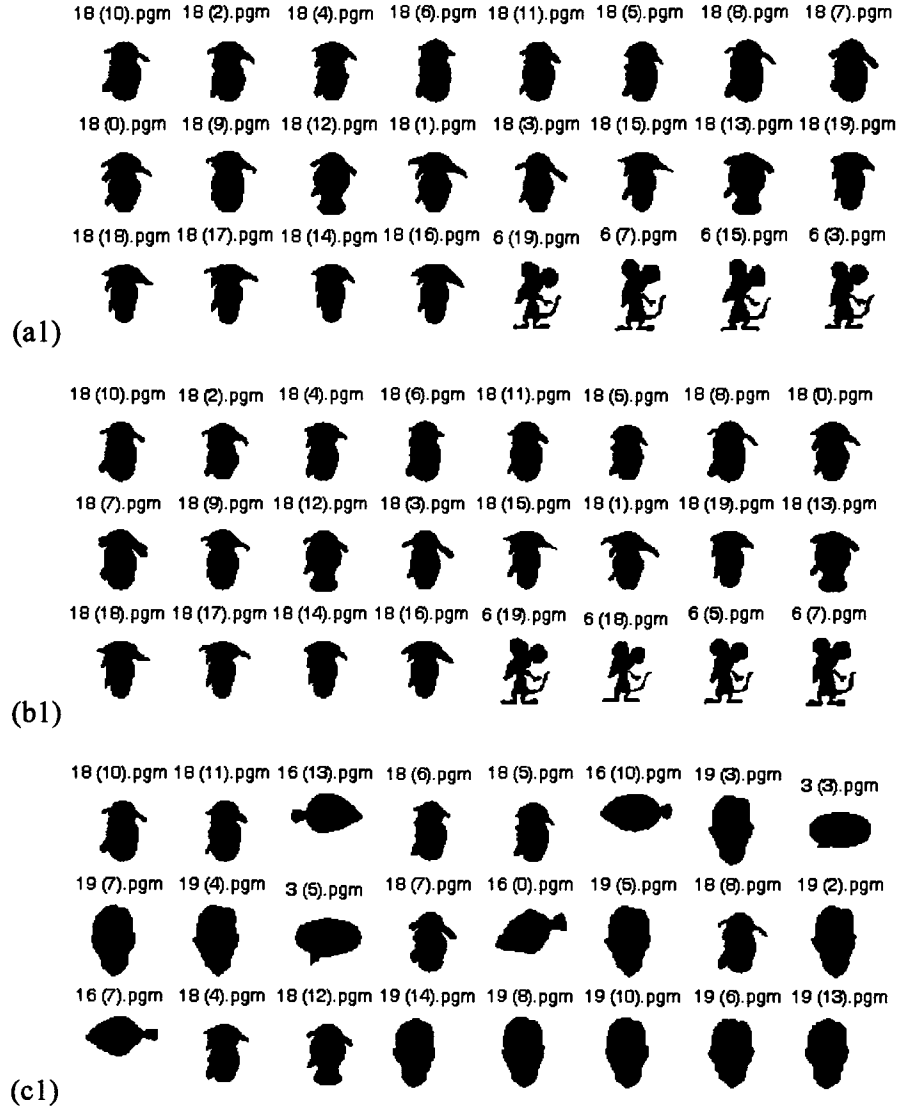
We choose performance measurement Precision and Recall Pair described in section 2.3.2. The statistic result, namely, the average precision and recall of all the 30 query images, is calculated from each subject. Finally, the average precision and recall of all the subjects is calculated and plotted in Figure 4.7.

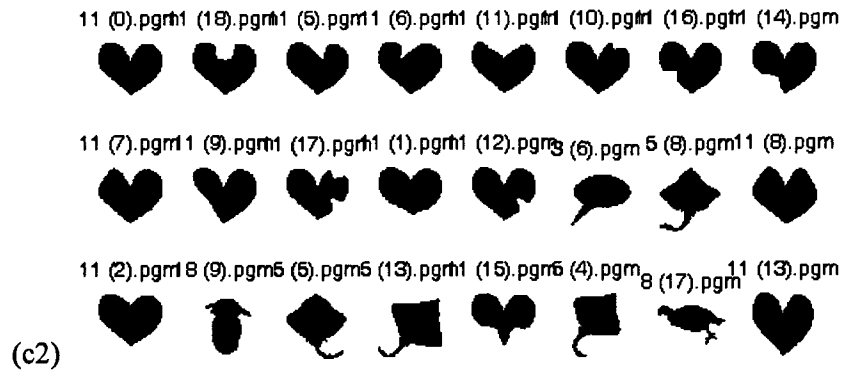
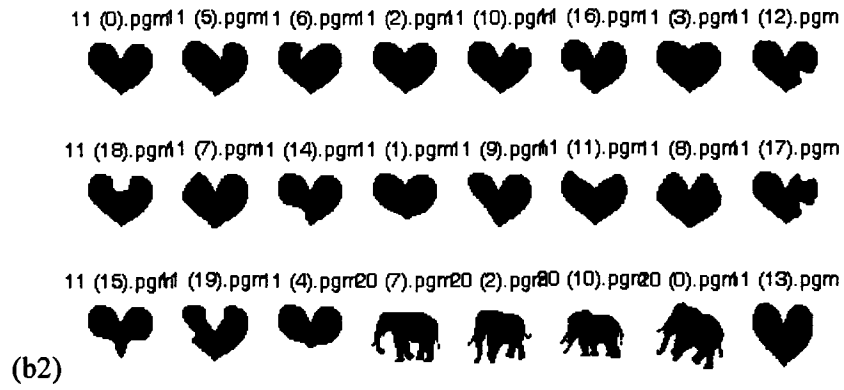
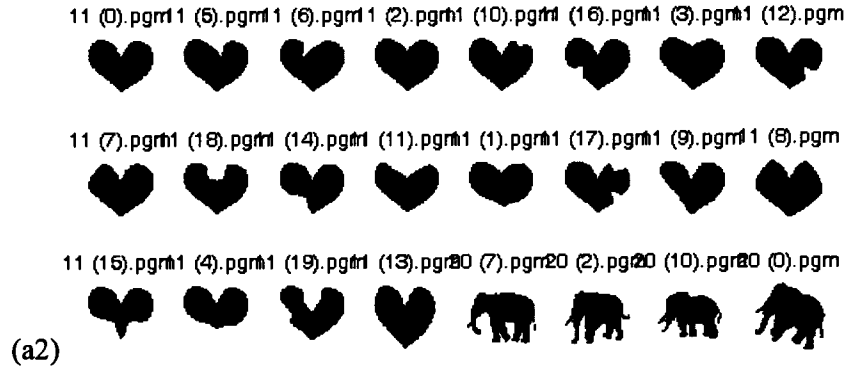


**Figure 4.7 Comparison of retrieval effectiveness of 5 FDs, 10 FDs and 15 FDs**

From figure 4.7, 5 FDs has the lowest retrieval performance, it is almost the same retrieval rate of 10FDs and 15FDs. The test reveals that 5 FDs has the lowest retrieval performance, when the number of FDs is above 10, the retrieval performance does not improve significantly with increased number of FDs to 15FDs, the retrieval performance does not decrease significantly when the number of FDs is reduced to 10 FDs. The results indicate that 10 FDs are sufficient for contour-based shape description with efficient retrieval rate. Although FD can retrieval most of similar images in database, it easily treat any shapes with hook-like parts as similar shapes to the query image. Like in 4.6 (a2) (b2) (c2), it retrieves the arms for marine fish because they have similar shapes.

Here we will compare the retrieval performance of GFD with FD. The example retrievals use 36GFD, 60GFDs and 10FDs is showed in figure 4.7, the similarity between features is measured by the Euclidean distance discussed in section 2.3.1. The first image is the query image in our experiment.

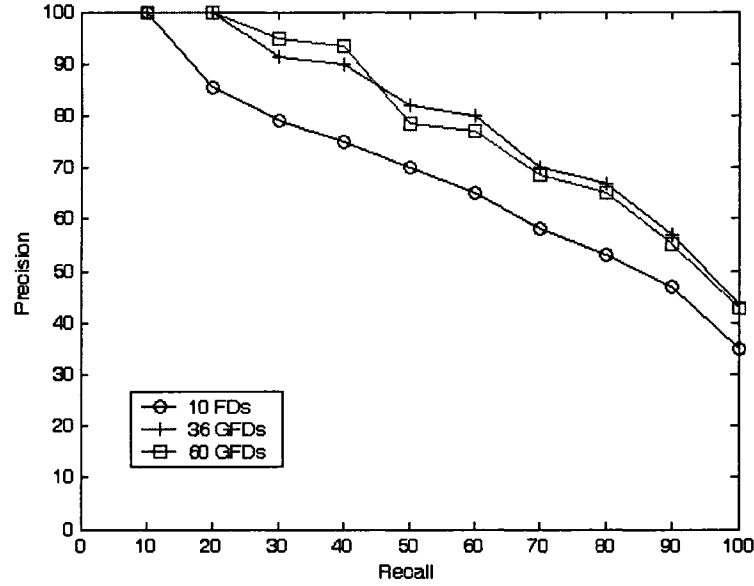




**Figure 4.8 Example retrievals using GFD and FD**

**(a1, a2) 36GFDs (b1, b2) 60GFDs (c1, c2) 10FDs**

We use precision and recall pair for retrieval evaluation. The statistic result, namely, the average precision and recall of all the 30 query images, is calculated from each subject for 36GFDs, 60GFDs and 10FDs respectively. Finally, the average precision and recall of all the subjects is calculated and plotted in Figure 4.9.



**Figure 4.9 Comparison of retrieval effectiveness of 10FDs, 36GFDs and 60GFDs**

From Figure 4.9, we can see clearly that GFD outperforms FD. The performance of FD is significantly lower than GFD. Since our database is a silhouette database, 36GFDs are enough to describe the contour information, so 36GFDs have higher retrieval rate than 60GFDs in overall performance. In this experiment, images retrieved by FD are confused with shapes with similar shape, but GFD overcomes this disadvantage by taking 2d Fourier transform of the whole image. GFD is also more robust to severe deformation of shape than the FD. Like in 4.8 (a2) (b2) (c2), GFD retrieves all of the heart images in the first 24 retrieval results while FD has lower performance due to some deformed or irregular heart images.

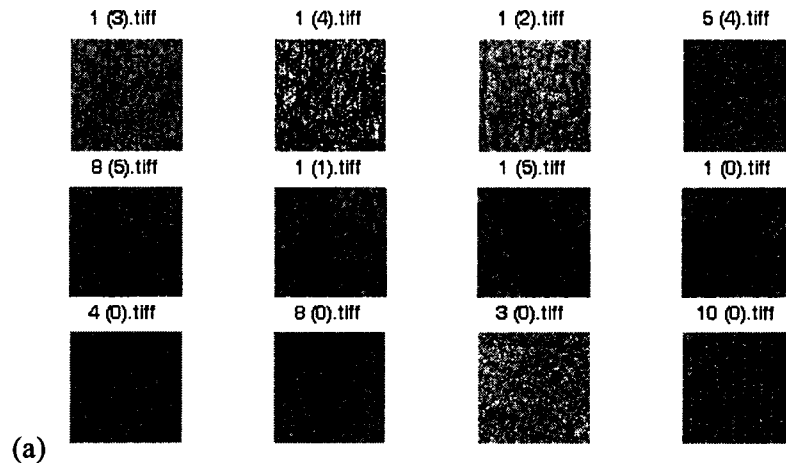
#### 4.4 Image Retrieval Using Texture Features

Texture is an important feature of object. When shape of object changes largely, its texture usually remains little changed. Therefore, if shape of object is complex, object's texture information can be exploited to describe object together with object's shape feature.

Gabor filters (GF) are a group of wavelets, with each wavelet capturing energy at a specific frequency and a specific direction. Expanding a signal using this basis provides a localized frequency description, therefore capturing local features/energy of the signal. Texture features can then be extracted from this group of energy distributions.

In this section, we will test the Gabor filters (GF) for texture feature extraction. GF also need parameter selection in (3.23) for the number of orientations and the number of scales. In our implementation, five scales and six orientations are used in common implementation for feature extraction.

The retrieval results are demonstrated on a Brodatz texture database. Our database consists of 20 different textures and each texture has 6 different images.



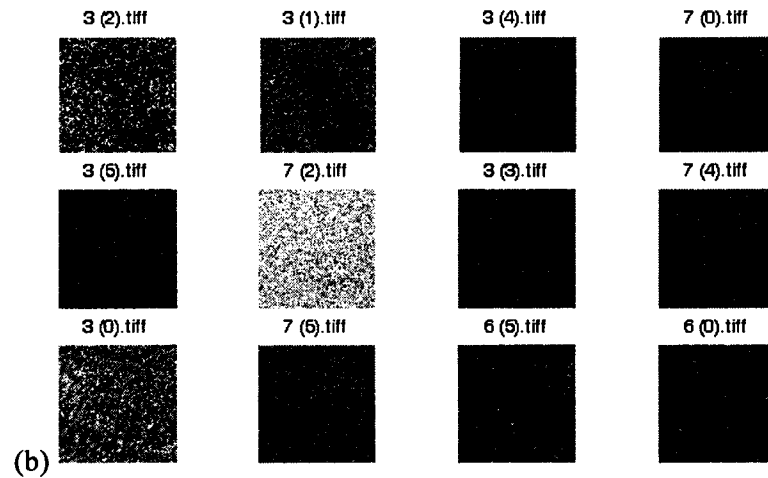


Figure 4.10 Example retrievals using GF

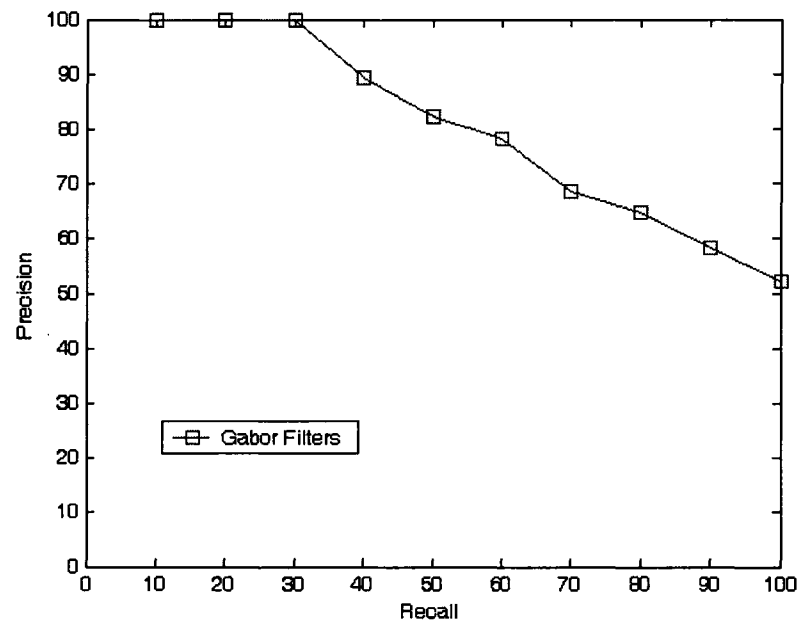


Figure 4.11 Retrieval Performance using Gabor Filters

The examples for retrieving texture images are showed in Figure 4.10 (a) and (b), the top left images are query image. From the retrieval results, two of the 6 similar texture images are retrieved in the first 9 images; we use 20 query images to calculate the

average precision and recall pair, which is plotted in Figure 4.11. The results show that Gabor texture features can efficiently represent texture images.

#### 4.5 Propose combined Shape and Texture Features

Previously, shape retrieval and texture retrieval have been tested on shape and texture image database individually. GFD has better performance than FD on shape-based image retrieval, and we achieve good results for texture image retrieval by using GF to obtain texture features. Shape is one of the most important low level image features due to that shape is a very important feature to human perception. Texture is a global feature that could be used to associate related shapes. When shape of object changes largely, its texture usually remains little changed. Intuition tells that combining object shape feature with texture feature gives more accurate description of object.

Based on Generic Fourier Descriptor and Gabor filters, we obtain GFD&GF feature vector for image retrieval. The overall distance between the query image and the database image is as follows:

(4.7)

Where  $d_{GFD}$  and  $d_{GF}$  are the distance measured by similarity measurements with features.  $w_{GFD}$  and  $w_{GF}$  are the weighting factors for respective parts of GFD and GF.

One way of choosing  $w_{GFD}$  and  $w_{GF}$  is to use relevance feedback for iterative setting [43]. The problem of relevance feedback is outside the scope of our research, we use  $w_{GFD} = w_{GF} = 0.5$  in our experiments.

In the next chapter, we will compare the retrieval performance of the proposed combined image features with that of individual image features like GFD and GF.

#### 4.6 Summary

In this chapter, a brief review of current segmentation techniques has been introduced at first. Then we make comparison and contrast on several feature extraction algorithms we discussed before. Two shape descriptors, Fourier Descriptor (FD) and Generic Fourier Descriptor (GFD), have been tested and compared on LEMS shape database. It shows that GFD has much better retrieval rate than FD. Gabor filters (GF) has been selected for texture feature extraction and it has good performance on a Fingerprint database. Based on the experiment results, we propose a combined shape and texture feature extraction algorithm at last.



## CHAPTER V

### EVALUATION AND EXPERIMENTAL RESULTS

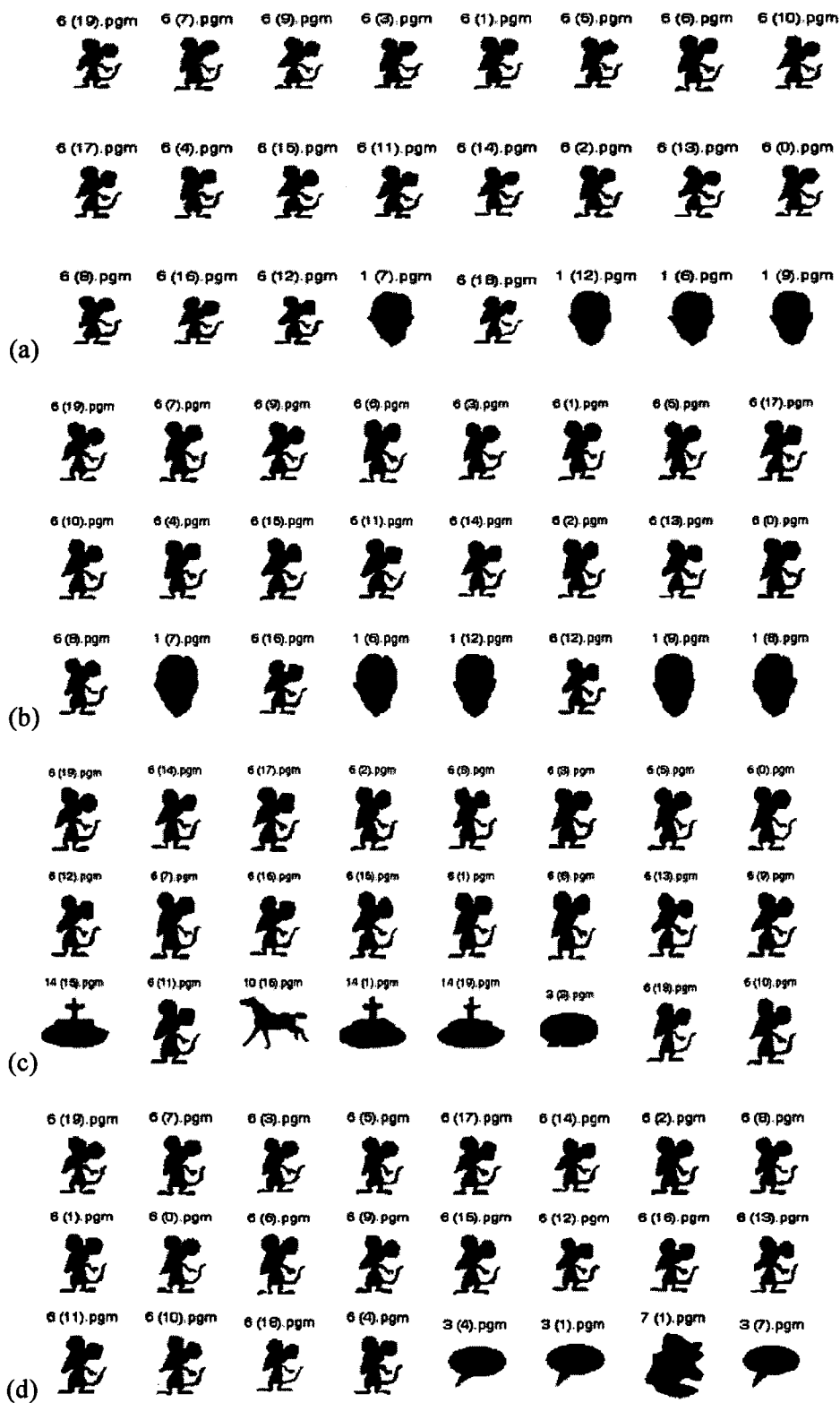
#### 5.1 Comparisons on Feature Extraction Algorithms

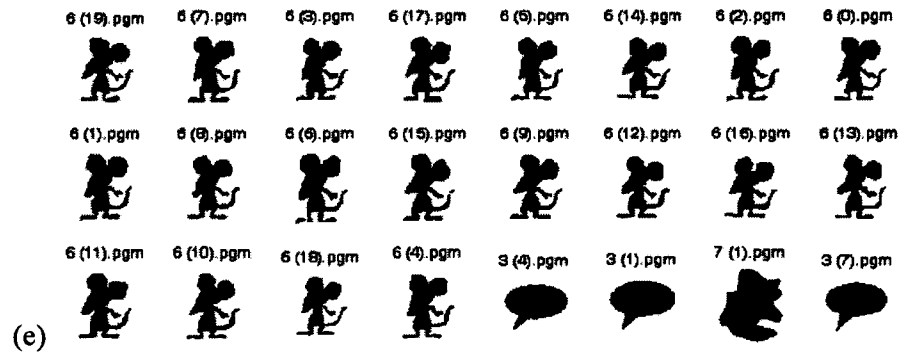
In order to test the combined features for robust and good retrieval performance, we compare the retrieval performance of combined image features with both GFD and GF for observation.

We conduct experiments on three different image databases: silhouette database collected by The Laboratory for Engineering Man/Machine Systems (LEMS), Fingerprint database and Amsterdam Library of Object Images (ALOI).

For parameter selection, we also choose 36GFDs and 60GFDs for Generic Fourier Descriptor, five scales and six orientations for Gabor filters. So our combined shape and texture features have two compositions: 36GFD+GF and 60GFD+GF. We will compare two combined features with 36GFD, 60GFD and GF.

The first database has been used in section 4.3 for evaluation FD and GFD. We have the same image databases here for comparison. It consists of 600 shape images with 30 subjects and 20 images per subject. 30 images (one image per class) are randomly selected as queries, which are shown in Figure 4.5. Here we use common similarity measure, the Euclidean distance to measure the similarity between two texture features. Figure 5.1 shows an example of retrieval results of all the feature extraction algorithms; the first image is the query image in our experiment.

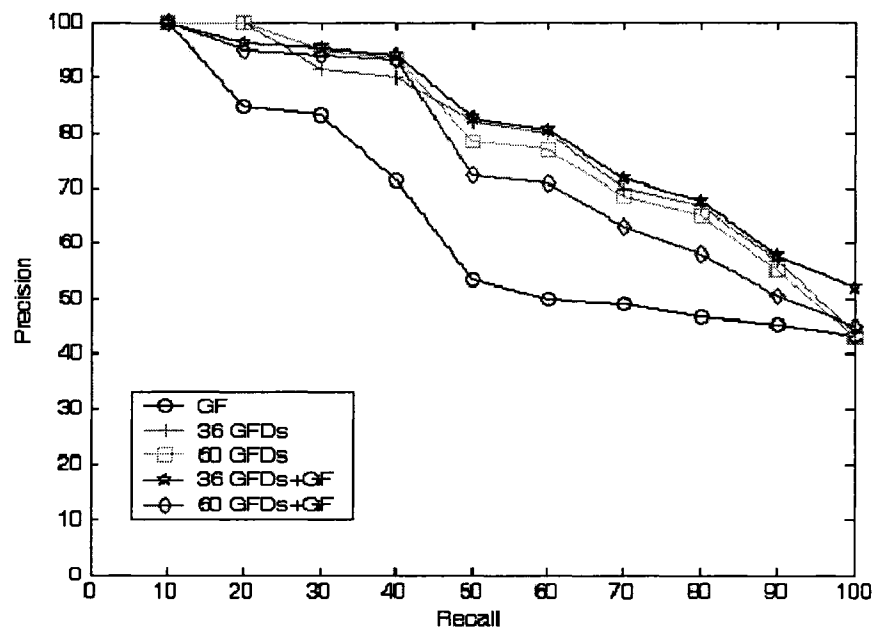




**Figure 5.1 Example retrievals using Combined features with GFD and FD**

(a) 60GFDs (b) 36GFDs (c) GF (d) 60GFDs+GF (e) 36GFDs+GF

Here we also choose performance measurement Precision and Recall Pair to evaluation and compare retrieval rate of different feature extraction algorithms. The precision and recall of all the 30 query images, is calculated from each subject. Finally, the average precision and recall of all the subjects is calculated and plotted in Figure 5.2.



**Figure 5.2 Comparison of retrieval effectiveness of GF, 36GFDs, 60GFDs, 36GFDs+GF, 60GFDs+GF**

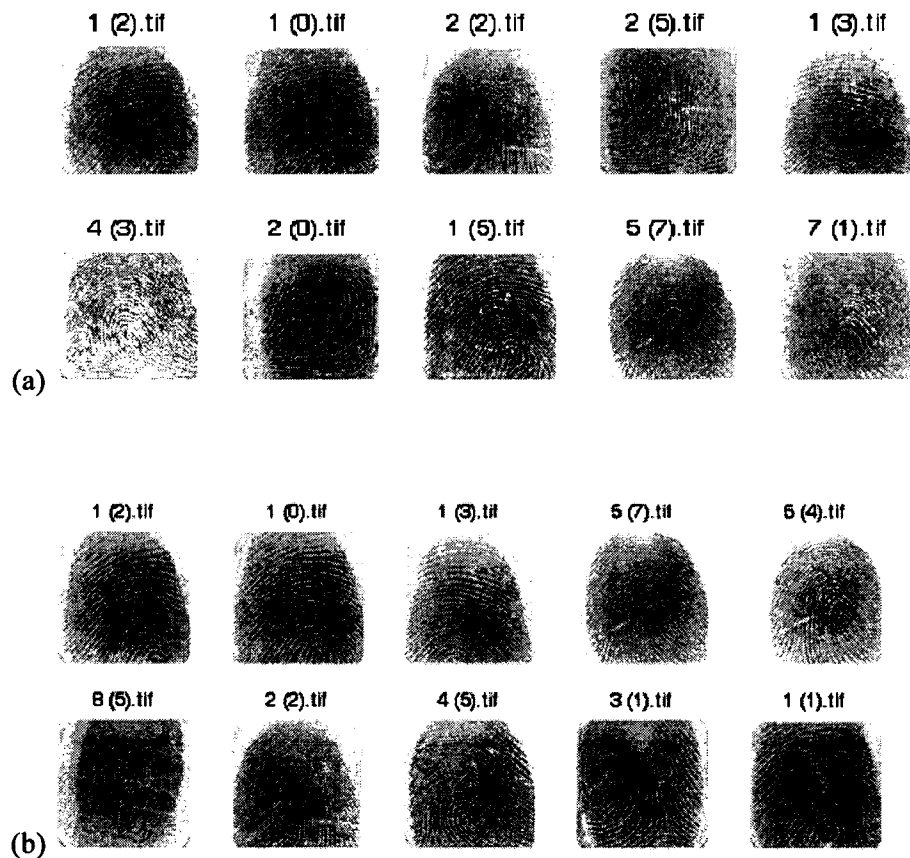
	Number of top matches				
Methods	1	5	10	15	20
<b>GF</b>	100	84.04	53.24	47.83	43.09
<b>36GFDs</b>	100	95.62	82.05	68.37	43.69
<b>60GFDs</b>	100	97.97	78.53	66.89	42.83
<b>36GFDs+GF</b>	<b>100</b>	<b>96.57</b>	<b>82.5</b>	<b>69.67</b>	<b>51.86</b>
<b>60GFDs+GF</b>	100	95.5	80.5	64.45	44.81

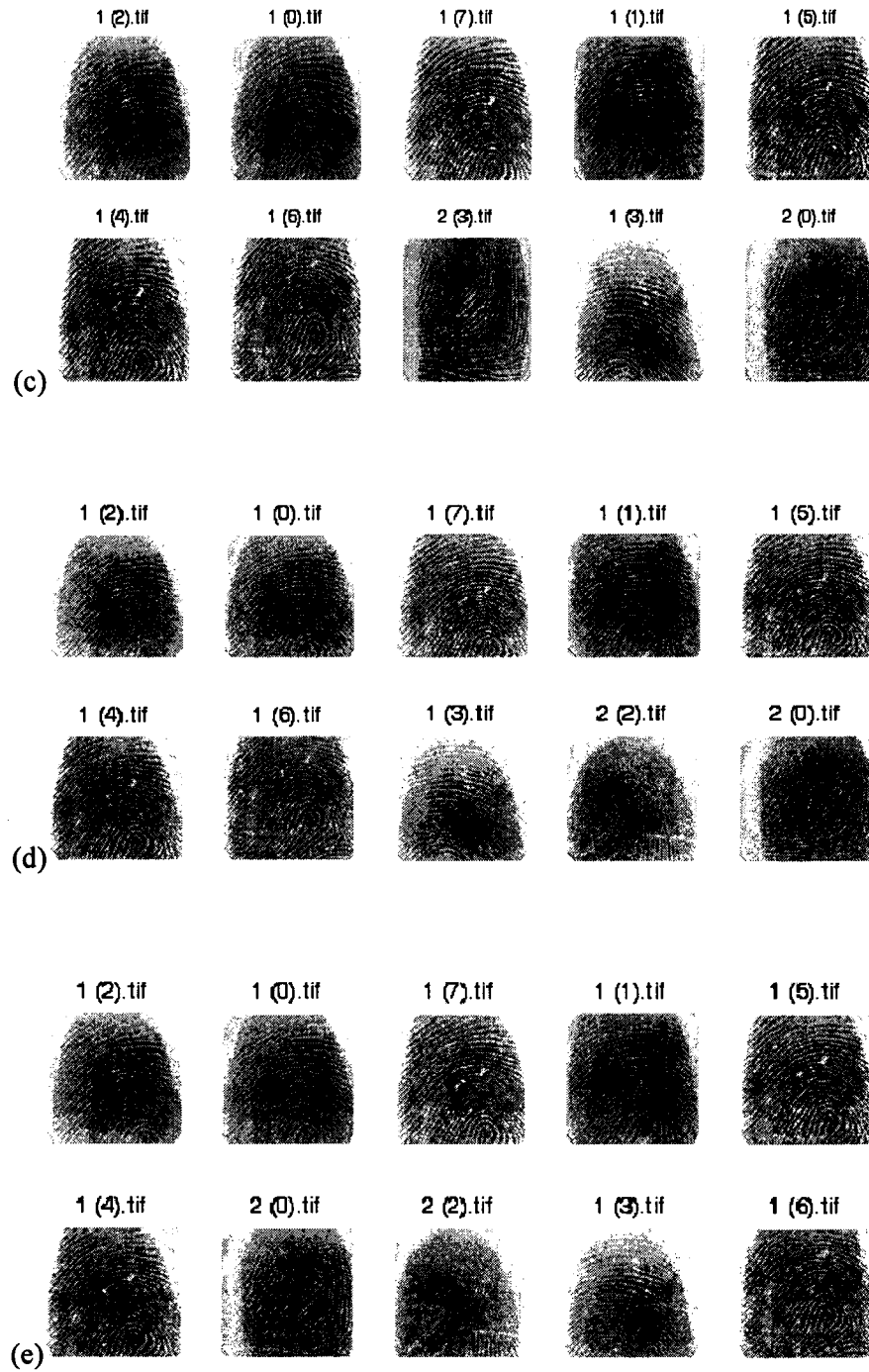
**Table 5.1 Retrieval Rate on LEMS silhouette database**

From Figure 5.2 that 36GFDs+GF achieves the best retrieval rate. GFD with 3 radial and 12 angular resolutions has a little higher performance than GFD with 4 radial and 15 angular resolutions. This might be the reason that our database is a silhouette database, 36GFDs are enough to describe the contour information, so 36GFDs have higher retrieval rate than 60GFDs in overall performance. The retrieval rates of the two GFDs are close to the best one 36GFDs+GF, while the GF's retrieval rate is much lower than others.

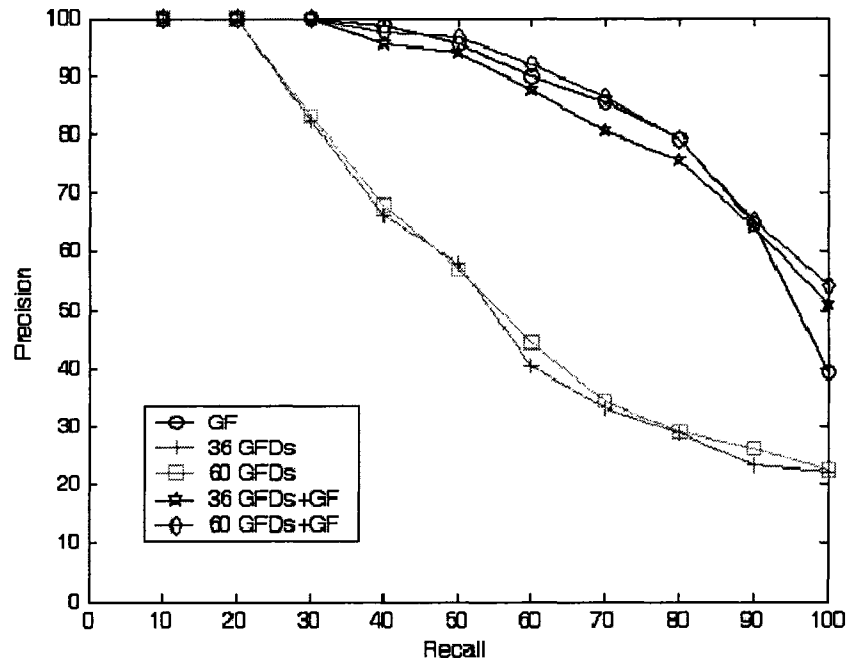
The results are showed clearly in table 5.1 that GFD and GFD+GF are similar in retrieval rate and 36GFDs+GF gives the highest retrieval rate. GF has the lowest performance since this is a shape database so GF can't effectively extract shape features. But combine GF with GFD, we can obtain better retrieval rate. So this experiment shows that the combined features are effective for shape-base image retrieval with best performance.

Our second experiments have been carried on a texture database, specifically, a Fingerprint database. This database is collected by FVC2000 [7]. It contains 110 fingers wide (w) and 8 impressions per finger. 50 images from different finger categories are randomly selected as queries. Here we use common similarity measure, the Euclidean distance to measure the similarity between two texture features. Figure 5.3 shows an example of retrieval results of all the feature extraction algorithms; the first image is the query image in our experiment. Figure 5.4 is the average precision and recall chart for all the subjects.





**Figure 5.3 Example retrievals using Combined features with GFD and FD**  
**(a) 60GFDs (b) 36GFDs (c) GF (d) 60GFDs+GF (e) 36GFDs+GF**



**Figure 5.4 Comparison of retrieval effectiveness of GF, 36GFDs, 60GFDs, 36GFDs+GF, 60GFDs+GF**

Methods	Number of top matches				
	1	2	4	6	8
<b>GF</b>	100	100	95.36	82.25	39.23
<b>36GFDs</b>	100	91.23	57.23	31.07	22.22
<b>60GFDs</b>	100	91.62	58.01	31.77	22.27
<b>36GFDs+GF</b>	100	100	93.9	77.98	51.12
<b>60GFDs+GF</b>	100	100	96.54	82.72	54.24

**Table 5.2 Retrieval Rate on Fingerprint database**

It shows from Figure 5.4 that 60GFDs+GF achieves the best retrieval rate. GF gets very high retrieval rate for fingerprint database. There is little difference (2%) between GF and GFD+GF in the retrieval rate, but GFD gives very low retrieval rate. It also

shows in table 5.2 that the combined features outperform the other two feature descriptors, GFD is not effective for texture image retrieval since it is a shape descriptor, but combine GFD with GF, we can obtain higher retrieval rate than GF alone. The results indicate that GF and GFD+GF are much more effective for fingerprint database than GFD.

Our third experiments select the object database collected by Amsterdam Library of Object Images (ALOI) [8]. ALOI is a color image collection of one thousand small objects. The images are systematically varied from viewing angle, illumination angle, and illumination color for each object. ALOI image database samples are showed in Figure 5.5.

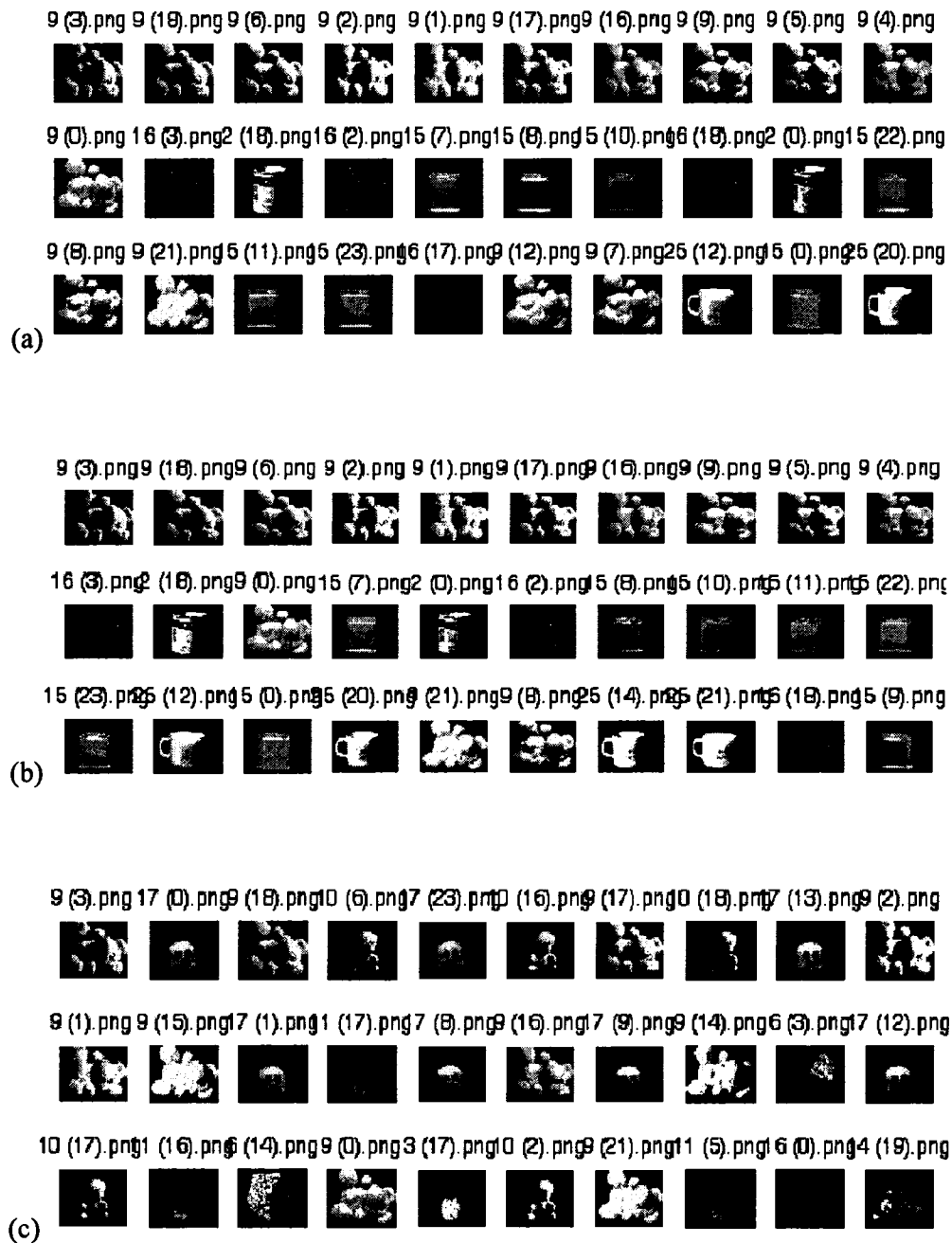


**Figure 5.5 ALOI image database samples [8]**

Our database consists 1200 images, and it's organized into 50 groups while 24 similar images in each group. In our experiment, we use gray level image for observation since we only extract shape and texture features of images. 50 images (one image per class) are randomly selected as queries. Here we use common similarity measure, the Euclidean distance to measure the similarity between two texture features.



Figure 5.6 shows an example of retrieval results of all the feature extraction algorithms; the first image is the query image in our experiment.



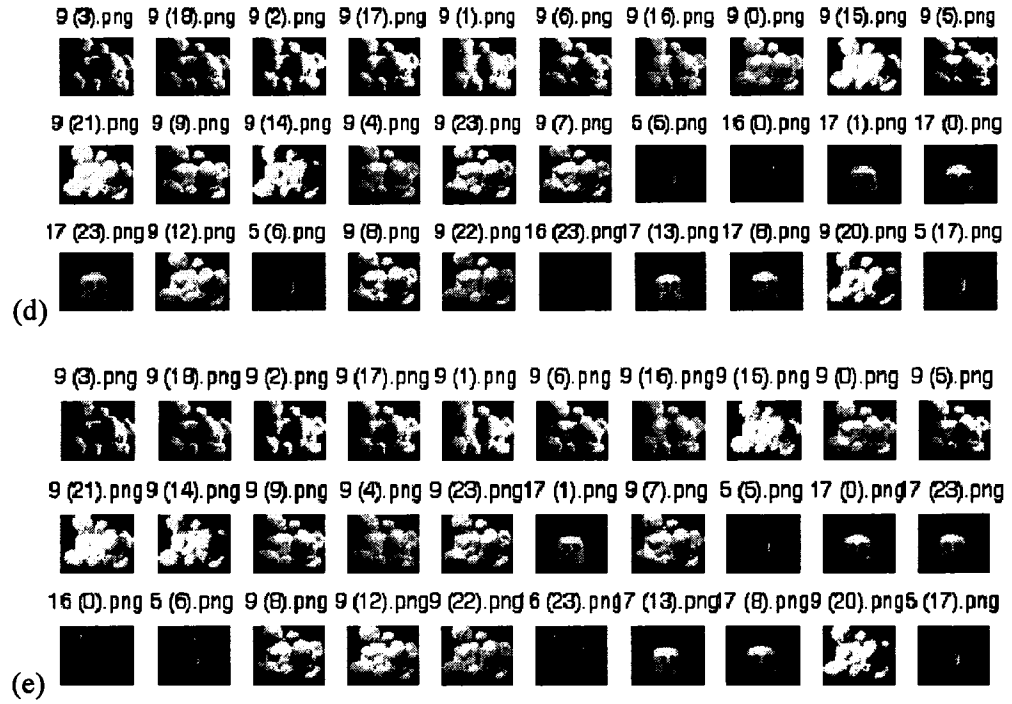


Figure 5.6 Example retrievals using Combined features with GFD and FD

(a) 60GFDs (b) 36GFDs (c) GF (d) 60GFDs+GF (e) 36GFDs+GF

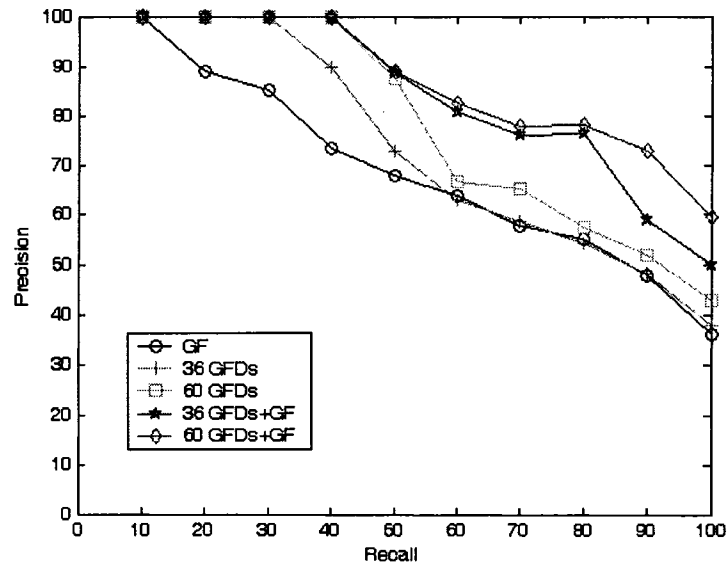


Figure 5.7 Comparison of retrieval effectiveness of GF, 36GFDs, 60GFDs, 36GFDs+GF, 60GFDs+GF

Here we also choose performance measurement Precision and Recall Pair to evaluation and compare retrieval rate of different feature extraction algorithms. The precision and recall of all the 50 query images, is calculated from each subject. Finally, the average precision and recall of all the subjects is calculated and plotted in Figure 5.7.

<b>Methods</b>	<b>Number of top matches</b>				
	1	4	12	20	24
<b>GF</b>	100	89.09	67.84	55.18	36.05
<b>36GFDs</b>	100	100	73.63	54.3	38.11
<b>60GFDs</b>	100	100	87.6	57.5	42.8
<b>36GFDs+GF</b>	100	100	88.64	76.52	49.77
<b>60GFDs+GF</b>	<b>100</b>	<b>100</b>	<b>89.02</b>	<b>78.10</b>	<b>59.55</b>

**Table 5.3 Retrieval Rate on ALOI database**

It can be seen from Figure 5.7 that 60GFDs+GF outperforms GFD and GF on ALOI database. Both of 36GFDs and 60GFDs achieve high retrieval rate, while GF has the lowest performance. In table 5.3, it shows 60GFDs has higher performance (average 7%) than 36GFDs. Although GF has the lowest retrieval rate, the overall performance is still good and it significantly improves the retrieval rate (average 13%) when combine GF and GFD together for image retrieval. The results indicate that both GFD and GF are effective for ALOI database and the combination of GFD and GF achieve the highest retrieval rate and effectively improve the overall performance.

## 5.2 Similarity Measurements

In this section, we will compare some common used similarity measurements indicated in 2.3.1, namely city block distance, Euclidean distance and  $\chi^2$  Statistics. In order to determine which distance measure best conforms to human perception, the retrieval effectiveness, and different distance measures is tested. Efficiency of the different similarity measurements is also tested. The similarity measurements are tested on two datasets used in last section.

The first datasets is silhouette database collected by The Laboratory for Engineering Man/Machine Systems (LEMS), which consists of 600 shape images with 30 subjects and 20 images per subject. 30 images (one image per class) are used as queries to test the retrieval result.

The second dataset is object image datasets collected by Amsterdam Library of Object Images (ALOI), which consists of 1200 images, and it's organized into 50 groups while 24 similar images in each group. 50 images (one image per class) are randomly selected as queries to test the retrieval result.

The feature describing the shape images in the first dataset is the Generic Fourier Descriptor (GFD), which is a normalized 36-dimension feature vector. The feature describing the images in the second database is the 64GFDs+GF. The retrieval performance measure used is the precision and recall. The precision and recall of all the 30 query images in silhouette database, is calculated from each subject. Finally, the average precision and recall of all the subjects is calculated and plotted in Figure 5.8.

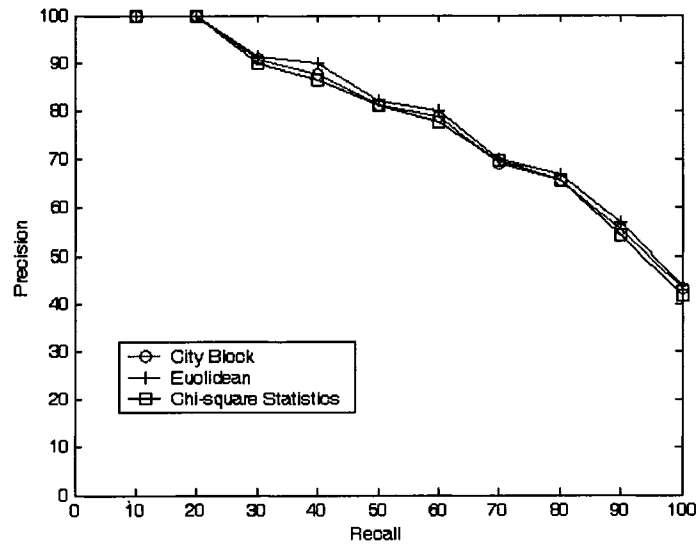


Figure 5.8 Retrieval performance of different similarity measurements on LEMS Dataset

Similarity Measure	Number of top matches				
	1	5	10	15	20
City Block	100	95.35	81.1	67.45	43.05
Euclidean	100	95.62	82.05	68.37	43.69
$\chi^2$ Statistics	100	95.23	81.0	67.7	41.8

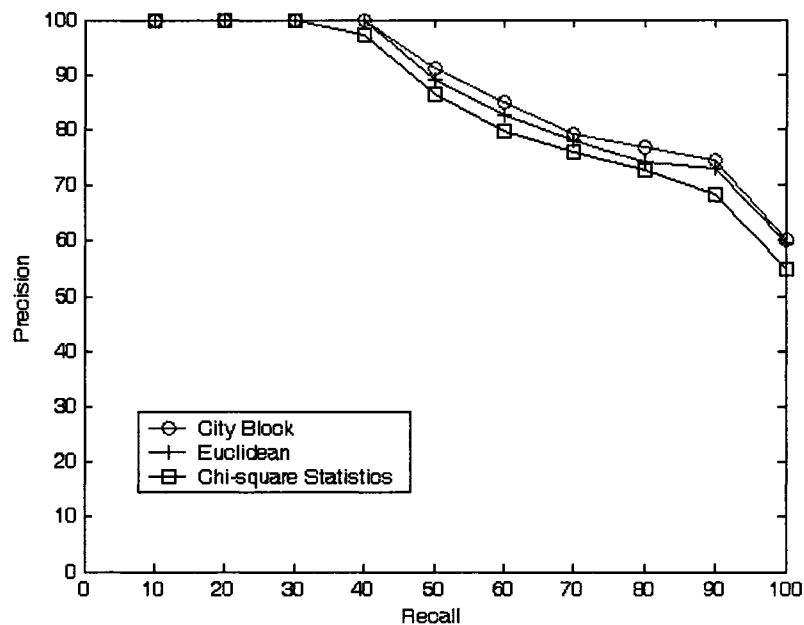
Table 5.4 Retrieval Rate on LEMS silhouette database of different similarity measurements

Similarity Measure	Total time taken for retrieving 600 shapes (s)	Time taken for retrieving each shape (ms)
City Block	37.5	62.5
Euclidean	43.2	70
$\chi^2$ Statistics	39.5	65.8

Table 5.5 Time taken for the retrieval on LEMS silhouette database

It is clearly showed in Figure 5.8 and Table 5.4 that city block distance, Euclidean distance, and  $\chi^2$  statistics have similar retrieval performance in shape database. Although Euclidean distance has highest retrieval rate, there is little difference ( 2%) performance among the three similarity measurements. Table 5.5 shows the time taken for retrieval all the 600 shape images; the time reported is obtained on Windows platform of a PC with Pentium M 1.6G processor and 512M memory.

The precision-recall obtained from the second dataset using the 3 similarity measurements is shown in Figure 5.9. The Retrieval Rate is calculated in Table 5.6, and time taken for the retrieval in ALOI dataset is given in table 5.7.



**Figure 5.9 Retrieval performance of different similarity measurements on ALOI Dataset**

Similarity Measure	Number of top matches				
	1	4	12	20	24
<b>City Block</b>	<b>100</b>	<b>100</b>	<b>91.1</b>	<b>75.55</b>	<b>60.05</b>
<b>Euclidean</b>	100	100	89.02	73.52	59.55
<b><math>\chi^2</math> Statistics</b>	100	100	86.5	70.45	54.8

**Table 5.6 Retrieval Rate on ALOI Dataset of different similarity measurements**

Similarity Measure	Total time taken for retrieving 1200 images (s)	Time taken for retrieving each image (ms)
<b>City Block</b>	98.6	82.2
<b>Euclidean</b>	102.6	85.5
<b><math>\chi^2</math> Statistics</b>	101.4	84.5

**Table 5.7 Time taken for the retrieval on ALOI Dataset**

From Figure 5.8 and Table 5.6, the city block distance achieves highest retrieval rate on ALOI image database while it has average 1.5% higher performance than Euclidean distance and average 5% than  $\chi^2$  statistics. The Table 5.7 indicates again that city block distance acquire the least time to retrieve images.

In summary, city block distance, Euclidean distance, and  $\chi^2$  statistics are all desirable distance measures in terms of both retrieval effectiveness and retrieval efficiency. However, city block distance is simpler than Euclidean distance and  $\chi^2$  statistics distance, it is the most desirable distance measurement for online retrieval.

### 5.3. Summary

In this chapter, the proposed combined features for image retrieval has been applied on three different image databases: shape database, fingerprint database and object database. And it has been compared with 36GFDs, 64GFDs and GF in each database. Results show that retrieval using combined shape and texture features is more powerful than retrieval using individual shape or texture features. Since our proposed features have been tested on both shape and texture databases, we can come to a conclusion that the proposed features can be used as a robust feature of shape and texture for image retrieval with highest performance.

Three similarity measurements have also been tested and evaluated in this chapter. The different similarity measurements have been evaluated using general image features and shape databases and object databases. Results show city block distance, Euclidean distance, and  $\chi^2$  statistics are all desirable distance measures in terms of both retrieval effectiveness and retrieval efficiency. But city block distance is more suitable for online retrieval due to its simple formulation and good retrieval rate.



## CHAPTER VI

### CONCLUSIONS AND FUTURE WORK

#### 6.1 Conclusions

In this thesis, techniques on content-based image retrieval (CBIR) have been investigated. Generally, CBIR involves three main issues—Image Pre-processing, Feature Extraction and Query Engine. In this research, feature extraction is the main interest.

Generally, in Content-based Image Retrieval, images consist three main contents: color, texture and shape, and images can be represented by the extracted features of image contents. Shape is the one of the most important features due to that shape is a very important feature to human perception. State of the art shape representation techniques have been reviewed. Shape-based image retrieval has been discussed. Two shape representation methods, region based shape descriptors and contour based shape descriptors have been comprehensively studied and evaluated. Although region-based shape representations can be applied to more general situations, they usually involve more computation. On the other hand, it is known that contours are so dominant in visual perception and other area. We describe the most favourite contour-based shape descriptor, Fourier descriptor (FD) in detail. Although FD has its own advantages and good performance for image retrieval, it has drawbacks. In order to overcome the disadvantages of the existing shape descriptors, a new shape descriptor, Generic Fourier Descriptor (GFD) has been discussed. We test FD and GFD on a shape database; results show that GFD performs well when outlines of objects are clearly defined and when

objects are not distorted beyond shape description. It has capability to retrieve partial shapes.

Texture feature extraction is another area in our research. Texture is also an important feature of object. Texture is a global feature that could be used to associate related shapes. When shape of object changes largely, its texture usually remains little changes. Therefore, if shape of object is complex, object's texture information can be exploited to describe object together with object's shape feature. So texture features could be a good supplement for shape feature extraction in image retrieval. Gabor filters (GF) has been selected for texture feature extraction and it shows good performance on a Fingerprint database.

Based on previous experiments and experience, we propose a combined shape and texture feature extraction algorithm for image retrieval. We test it on three different image databases: shape database, fingerprint database and object database. And it has been compared with 36GFDs, 64GFDs and GF in each database. Results show that on overall, retrieval using combined shape and texture features is more powerful than retrieval using individual shape or texture features. Since our proposed features have been tested on both shape and texture databases, we can come to a conclusion that the proposed features can be used as a robust feature of shape and texture for image retrieval with highest performance.

Similarity measurements have also been evaluated in this research. Results show city block distance, Euclidean distance, and  $\chi^2$  statistics are all desirable distance measures in terms of both retrieval effectiveness and retrieval efficiency. But city block

distance is more suitable for online retrieval due to its less retrieval time and relatively good performance.

## 6.2 Future Works

Content-based image retrieval is a difficult task. Although much work has been done in this area, many issues are still unsolved and need more research. Many advances have been made in various research aspects, including visual content description, similarity/distance measures, indexing scheme, user interaction and system performance evaluation. The most outstanding issue is to describe image using higher-level features. However, the low-level features do not have a direct link to the high-level concepts. It is an important research area in CBIR to narrow down this semantic gap through off-line and on-line processing.

User interaction is also an important issue in CBIR. On one hand, images to be retrieved are determined by the user's specification of the query. On the other hand, query results can be refined to include more relevant candidates through the relevance feedback of users. Updating the retrieval results based on the user's feedback can be achieved by updating the images, the feature models, the weights of features in similarity distance, and select different similarity measures. Relevance feedback from user will be looked into in future research.

## REFERENCES

- [1] H. Tamura and N. Yokoya, *Image database systems: A survey*, *Pattern Recognition* 17(1), 1984
- [2] S.-K. Chang and A. Hsu, *Image information systems: Where do we go from here?* IEEE Trans. on Knowledge and Data Engineering 4(5), 1992.
- [3] Y. Rui, T.S. Huang, and S.F. Chang, *Image Retrieval: Current Techniques, Promising Directions, and Open Issues*, Journal of Visual Communication and Image Representation, Volume 10, pp. 39-62, 1999
- [4] Long F., Zhang H.J., and Feng D.D. *Fundamentals of Content-Based Image Retrieval" in Multimedia Information Retrieval and Management-Technological Fundamentals and Applications*, D. Feng, W.C. Siu, and H.J. Zhang (Eds.), Springer, 2003
- [5] B.S.Manjunath and W.Y. Ma: *Texture features for browsing and retrieval of image data*, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol.18, no.8, pp.837-42, 1996
- [6] LEMS Vision Group, <http://www.lems.brown.edu/vision/software/index.html>
- [7] FVC2000, <http://bias.csr.unibo.it/fvc2000/>
- [8] Amsterdam Library of Object Images (ALOI), <http://staff.science.uva.nl/~aloi>
- [9] J. Wang, W.J. Yang, and R. Acharya, *Color clustering techniques for color-content-based image retrieval from image databases*, in Proc. IEEE Conf. on Multimedia Computing and Systems, 1997

- [10] M. Swain and D. Ballard, *Color indexing*, International Journal of Computer Vision 7(1), 1991.
- [11] M. Stricker and M. Orengo, *Similarity of color images*, in *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1995.
- [12] M. Stricker, A. Dimai, *Color indexing with weak spatial constraints*. Proc. SPIE. Storage Retrieval Still Image Video Databases, IV 2670, 1996
- [13] C. Carson, S. Belongie, H. Greenspan, and J. Malik, Region-based image querying, in Proc. of IEEE Workshop on Content-Based Access of Image and Video Libraries, in Conjunction with IEEE CVPR'97, 1997
- [14] M. Stricker and M. Orengo, *Similarity of color images*, in Proc. SPIE Storage and Retrieval for Image and Video Databases, 1995.
- [15] R. M. Haralick, K. Shanmugam, and I. Dinstein, *Texture features for image classification*, IEEE Trans. On Sys. Man. and Cyb. SMC-3(6), 1973.
- [16] H. Tamura, S. Mori, and T. Yamawaki, *Texture features corresponding to visual perception*, IEEE Trans. on Sys., Man. and Cyb. SMC-8(6), 1978.
- [17] J. R. Smith and S. F. Chang, *Transform features for texture classification and discrimination in large image databases*, in Proc. IEEE Int. Conf. on Image Proc., 1994.
- [18] J. R. Smith and S.-F. Chang, *Automated binary texture feature sets for image retrieval*, in Proc. ICASSP-96, Atlanta, GA, 1996.
- [19] C. T. Zahn and R. Z. Roskies, *Fourier descriptors for plane closed curves*, IEEE Trans. on Computers, 1972.

- [20] E. Persoon and K. S. Fu, *Shape discrimination using fourier descriptors*. IEEE Trans. Sys. Man. Cyb, 1977.
- [21] Gene C.-H. Chuang and C.-C. Jay Kuo. *Wavelet descriptor of planar curves: Theory and applications*. IEEE Trans. Image Proc., 5(1), pp 56-70, 1996
- [22] Abbasi, S., Mokhtarian, F., Kittler, J. *Curvature scale space image in shape similarity retrieval*. Multimedia Systems 7 (6), pp 467 – 476, 1999.
- [23] C.-H. Teh, R.T. Chin, *On image analysis by the methods of moments*, IEEE Trans. Pattern Anal. Mach. Intell. 10 (4), pp 496 – 513, 1988
- [24] G.J. Lu, A. Sajjanhar, *Region-based shape representation and similarity measure suitable for content-based image retrieval*, Multimedia Syst. 7 (2), 1999.
- [25] H. Voorhees and T. Poggio. *Computing Texture Boundaries from Images*. Nature, 333, pp 364-367, 1988.
- [26] Y. Rubner. *Perceptual Metrics for Image Database Navigation*. PhD thesis, Stanford University, 1999.
- [27] G. Lu and A. Sajjanhar. *On Performance Measurement of Multimedia Information Retrieval Systems*. In Proc. of International Conference on Computational Intelligence and Multimedia Applications, pp.781-787, Australia, Feb. 9-11, 1998.
- [28] A. Del Bimbo. *Visual Information Retrieval*. Morgan Kaufmann Publishers, Inc. San Francisco, USA, pp. 56-57, 1999.

- [29] M. Bober, *MPEG-7 Visual Shape Descriptors*, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 11, NO. 6, JUNE 2001
- [30] F. L. Bookstein. *Principal warps: Thin-plate splines and the decomposition of deformations*. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-11(6), pp 567-585, 1989.
- [31] D. Zhang and G. Lu, *A Comparative Study of Curvature Scale Space and Fourier Descriptors for Shape-Based image Retrieval*, Journal of Visual Communication and Image Representation, pp. 41-60, Vol. 14, No.1, 2003
- [32] D. Zhang, G. Lu: *Evaluation of MPEG-7 shape descriptors against other shape descriptors*. Multimedia Syst. 9(1), pp 15-30, 2003
- [33] D.S. Zhang and G. Lu, *Shape-based image retrieval using generic Fourier descriptor*, SP:IC(17), pp. 825-848, No. 10, November 2002
- [34] D. Zhang and G. Lu. *Content-Based Shape Retrieval Using Different Shape Descriptors: A comparative Study*. In Proc. of IEEE Conference on Multimedia and Expo (ICME'01), Tokyo, pp.317-320, August 2001.
- [35] F. Liu and R. W. Picard: *Periodicity, directionality, and randomness: World features for image modeling and retrieval*. Technical report, MIT Media Laboratory and Modeling Group, technical Report No. 320, 1994.
- [36] R. Jain, R. Kasturi and B. G. Schunck. *Machine Vision*. McGraw-Hill Inc. 1995.
- [37] D. Marr and E. Hildreth. *Theory of Edge Detection*. Proc. Roy. Soc. London, B207, pp.187-217, 1980.

- [38] J. Canny. *A Computational Approach to Edge Detection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6):679-697, 1986
- [39] M. Sonka, V. Hlavac, R. Boyle. *Image Processing, Analysis and Machine Vision*. Chapman & Hall Computing, 1993.
- [40] R. M. Haralick and L. G. Shapiro. *Survey: Image Segmentation Techniques*. Computer Vision, Graphics and Image Processing, 29:100-132, 1985.
- [41] S. Beucher. *The watershed transformation applied to image segmentation*. Conference on Signal and Image Processing in Microscopy and Microanalysis, pp 299-314, September 1991.
- [42] M. M. Reid, R. J. Millar and N. D. Black. *Second-Generation Image Coding: An Overview*. ACM Computing Surveys, 29(1), pp 3-29, 1997.
- [43] Kyoung-Mi Lee, W. Nick Street: *Incremental feature weight learning and its application to a shape-based query system*, Pattern Recognition Letters, v.23 n.7, pp 865-874, 2002



## VITA AUCTORIS

NAME:	Quan He
PLACE OF BIRTH:	Wuhan, People's Republic of China
EDUCATION:	Wuhan Univerity, Wuhan, China 1999-2003 Bachelor of Engineering  University of Windsor, Windsor, Ontario 2005-2007 Master of Applied Science